IDENTIFYING THE GENES AND TISSUES REGULATING BLOOD PRESSURE AND ITS VARIATION

by
Priyanka Nandakumar

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, MD
September, 2017

Abstract

Hypertension is a complex trait with approximately 20-50% heritability. Several genome-wide association studies (GWAS) have implicated hundreds of loci involved in blood pressure (BP) regulation, but these variants explain only a small proportion of the phenotypic variance, and the specific genes and variants at most of these loci have yet to be determined. Additionally, while 20 genes involved in renal salt homeostasis have been well-characterized in monogenic forms of hypertension or hypotension, the extent to which they contribute to primary or essential hypertension is unknown.

Hypertension is disproportionately prevalent in African Americans relative to other American populations, and yet African Americans are underrepresented in studies designed to advance genetic discoveries for this disease. The study described in Chapter 2 addresses this issue by examining the genotypes of 15,914 African-ancestry individuals for rare and low-frequency variants, to further BP gene discovery and to explain more of the phenotypic variance of BP traits in African Americans. We identified rare variants in 10 genes, with many supported by previous functional evidence of cardiovascular and related roles.

In Chapter 3, we examined the genes *SLC12A3*, *SLC12A1*, and *KCNJ1* in an effort to replicate the BP-lowering effect of predicted loss-of-function variants as reported in a prior Framingham Heart Study (FHS) publication. While analysis of all variants meeting the FHS criteria in exome sequencing data from a second cohort, the Atherosclerosis Risk in Communities (ARIC) study, did not produce a BP-lowering effect, we did observe a similar reduction among carriers (~6 mmHg for SBP, ~3 mmHg

for DBP) of a subset of 10 variants from the FHS study also appearing in the ARIC data, suggesting that variants at specific sites may be of interest within these genes.

We shifted the focus from exome studies to whole genome analysis of putative regulatory variation in Chapter 4, where we first identified tissues relevant to BP regulation and subsequently followed up with gene and variant discovery incorporating tissue-specific information in the analysis. We were able to identify several genes with strong expression support in the tissues of interest and for some genes, identify their tissue- or cell-type-specificity.

Thesis readers:

Dr. Aravinda Chakravarti, PhD (advisor)

Dr. Priya Duggal, PhD, MPH

Preface

I am grateful to my advisor, Dr. Aravinda Chakravarti, for his mentorship and support through the experiences and learning opportunities I have had working in his lab. The members of my thesis committee, Drs. Steven Salzberg, Josef Coresh, Victor Velculescu, and Hans Bjornsson, have also provided constructive feedback on the work described in this thesis over these last few years. I would also like to thank Dr. Priya Duggal for assisting in the completion of the Wolfe Street Competition RNA sequencing project, and for serving as one of my thesis readers.

There are many collaborators whose contributions have made the work described in this thesis possible. They include: Eric Boerwinkle, Alanna Morrison, Megan Grove, Myriam Fornage, and Melissa Richard at the University of Texas Health Science Center at Houston; C. Charles Gu and Yun Ju Sung at Washington University in St. Louis; Sharon Kardia and Erin Ware at the University of Michigan; Richard Cooper and Bamidele Tayo at Loyola University, Chicago; Charles Rotimi and Fasil Tekola-Ayele at the National Institutes of Health; Georg Ehret at Geneva University Hospital/JHMI; Neil Risch, Thomas Hoffmann, Pui-Yan Kwok, and Mark Kvale at the University of California, San Francisco; and Carlos Iribarren, Catherine Schaefer, and Dilrini Ranatunga at Kaiser Permanente. I am also thankful to all of the members of the lab, and have had the privilege of working with several of them on various projects.

The time I have spent in the Human Genetics program has been valuable, for which I would like to thank Dr. David Valle, Sandy Muscelli, the program executive committee members, as well as many of the other students and members of the department that I have enjoyed meeting and getting to know over the last several years.

My ultimate appreciation goes to my family and friends, for their long-standing support and encouragement, especially through the last several years.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

What is hypertension?

Hypertension, or chronically elevated blood pressure, is defined as systolic blood pressure (SBP) ≥ 140 mmHg or diastolic blood pressure (DBP) ≥ 90 mmHg; clinically, the ideal blood pressure is defined as 115/75 (SBP/DBP) mmHg.[1] Hypertension per se is not a disease but its is a major risk factor for cardiovascular disease, and associated diseases such as kidney disease and diabetes.[1] The prevalence of hypertension in the United States is ~30%, frequently unrecognized, and is a condition with an incidence disproportionately elevated in African Americans as compared to other Americans.[1]

Various forms of anti-hypertensive therapies have been developed to reduce the risk of hypertensive disease,[2] but only 53% of such individuals have effective blood pressure (BP) control. The majority of these individuals is non-Hispanic white, with lower frequencies of controlled hypertension among ethnic minority groups, which may be due a combination of genetics and socioeconomic factors.[3] The studies described in this thesis seek to advance our understanding of blood pressure genetics to improve our understanding of its' disease mechanisms, and assist in informing therapies as well as public health approaches to address issues of high prevalence and insufficient treatment of hypertension (HTN).

Genetics of blood pressure

Primary hypertension is postulated to result from polygenic inheritance, with estimates of approximately 20-50% heritability.[4–6] But, what about its mechanisms?

Approximately 20 genes with well-established roles in monogenic, syndromic forms of hypertension or hypotension involve renal salt homeostasis.[7–11] This suggests its similar role in primary (aka, essential) hypertension, but this remains unproven. Consider that over two dozen large-scale genome-wide association studies (GWAS), in addition to several other studies, have identified common variants at hundreds of loci, residing mainly in non-protein-coding regions, that are associated with blood pressure (BP) traits,[12–36] but that these contain just one of the monogenic blood pressure syndrome genes, *CYP17A1*.[13,28] These studies do confirm the polygenic nature of BP control, however, genetic risk score analysis of 29 loci in one such large-scale study demonstrated association of these BP GWAS variants with cardiac damage and cardiovascular disease,[12] but not with kidney disease or function. Further, most BP studies, of both monogenic genes and BP GWAS, have been conducted primarily in European-ancestry subjects, and just a small fraction of the BP phenotypic variance has been explained by the known genes and loci,[33,37–39] emphasizing what we do not know of the "missing heritability."

At this time, the relative contributions of the monogenic genes and genes at BP GWAS loci to primary hypertension remain to be further examined; further, the specific causal genes and variants at the BP loci are yet to be determined. These open questions have motivated the primary aims of this dissertation, which are to identify the specific candidate genes and variants involved in BP regulation, both at novel and previously known loci, and to examine the contributions of these genes to blood pressure. The initial studies I conducted analyze the exome (Chapters 2 and 3), with the focus then shifting to analysis of regulatory variation across the genome (Chapter 4). Chapter 2 aims to further

BP gene discovery in African Americans, while Chapter 3 examines rare variants in known BP monogenic genes to assess their effects. The aim of Chapter 4 is to begin to dissect associated variants at BP GWAS loci as specific, but as yet putative, regulatory variation in tissues of interest given their primary location in the non-protein-coding genome.

Identifying blood pressure genes in populations of interest

As already stated, the hundreds of GWAS BP-associated loci have mostly been identified from European-ancestry populations, and, additionally, only explain a small (~3%) percentage of the phenotypic variance. [33,39] Further, the higher prevalence of hypertension and associated diseases in African Americans, coupled with the comparatively less successful treatment in this population, calls for further genetic discovery in this population. The study described in Chapter 2 was designed to serve two purposes: 1) to advance blood pressure genetics in a population highly affected by the condition yet underrepresented in such studies, and 2) to identify rare and low-frequency variants that may explain more of the BP phenotype. We addressed these questions by genotyping African-ancestry individuals on the Illumina HumanExome BeadChip.[40] This exome array includes variants specifically selected for their putative functional significance and low frequency, and at a sample size of 15,914 individuals across eight studies, it is the second largest African-ancestry GWAS to date, and the largest to focus on exomic variants. We identified rare variants in 10 genes in this study, with supporting evidence for previously replicated blood pressure loci.

Expanding on studies of syndromic forms of blood pressure

The BP literature identifies 20 genes containing rare variants with large effect sizes (>6 mmHg)[7–11] involved in renal salt homeostasis and involved in monogenic forms of syndromic hypertension or hypotension. This has had great utility in identifying and establishing blood pressure drug targets, such as the multiple classes of diuretics to maintain renal salt and fluid balance.[41] Nevertheless, as mentioned, the phenotypic variance explained by these rare variants is also small.[37,38] While each of the main chapters includes some study of the monogenic genes, Chapter 3 outlines a study focused on replicating the previously observed effects of rare variants within the coding regions of three hypotension genes, *SLC12A3*, *SLC12A1*, and *KNCJ1*.[10] The purpose of doing so was to examine the effect of variants meeting the same criteria to be considered as predicted loss-of-function from the prior study and to identify additional variants of interest, using exome sequencing data from the Atherosclerosis Risk in Communities (ARIC) European-ancestry population-based cohort. We discovered that analysis of all variants meeting the functional criteria to be considered damaging did not produce the expected BP-lowering effect, but that studying the subset of overlapping variants between their study and ours did. Consequently, variants at specific sites may be of interest for blood pressure regulation but we may have failed to observe the expected effect due to the presence of sequencing or other errors in our study.

Additionally, as described in the next section, we aimed to identify putative regulatory variants specific to these monogenic genes. An expanded list of variants would allow a greater understanding of the various blood pressure pathways these genes are involved in beyond what is known, as well as their regulatory networks.

Mapping blood pressure loci in tissues of interest

With an ever-growing list of well-replicated loci from BP GWAS, the main goal of Chapter 4 was to begin to elucidate specific causal genes and variants at these loci and to begin to dissect some of the unknown biology of BP regulation. The role of the kidney in BP regulation has long been acknowledged, but a systemic trait like blood pressure involves many other tissues. In recent years, the growth of data in public genomics databases has now made it possible to use such functional information to assess the role of these 'other' (non-renal) tissues.

The Genotype-Tissue Expression (GTEx) project (https://www.gtexportal.org/) includes genotype and expression data from 53 different tissues, with expression quantatitive trait loci (eQTL) available for 44 of them, all having sample size > 70 individuals. These data allow us to examine sequence variant-gene expression correlation, on a tissue-by-tissue basis. The Roadmap Epigenomics project (http://www.roadmapepigenomics.org/) and the Encyclopedia of DNA Elements (ENCODE) project (https://www.encodeproject.org/) contain among other types of data, genotype, gene expression, chromatin accessibility and histone modification data from different tissues, time points, and organisms. The purpose of these projects is to integrate these different datasets to identify and determine context-specific functional genomic elements.

Standard analysis methods of genotype data involve testing variants at the variant level and at the gene-based level. Testing individual variants for BP association has been successful in large GWAS as described above, in which substantial sample sizes enable

identification of common variants with small effects. Gene-based tests, such as burden tests, such as T1 or T5, which analyze variants with minor allele frequencies < 1% or <5%, respectively,[42] as well as non-burden tests like SKAT,[43] which collapses and tests groups of variants, are usually used to detect the effects of a collection of low frequency and rare variants. Additionally, the development of functional scores for variants, such as CADD,[44] GWAVA,[45] PolyPhen,[46] and SIFT[47] scores, has allowed the selection of those at either the exome or genome levels that may be considered functionally damaging and to be prioritized for testing. However, these scores generally do not account for tissue or cell-type specificity. Therefore, analyses that incorporate or select on these scores in genome-wide analyses may identify genes of interest, but do not test genes in a tissue-specific manner.

The primary source of BP heritability has previously been identified as DNaseI hypersensitivity sites (DHSs).[6] In Chapter 4, we first focus on identifying tissues of interest for BP regulation by testing various tissues for enrichment of BP loci in eQTLs, using data from the GTEx project, revealing the aorta and tibial arteries as specific candidate tissues. We then focus on identifying the effects of putative cis-regulatory elements and on mapping BP loci in tissues of interest, incorporating deltaSVM scores,,[48] a new type of score that predicts the impact of variants on regulatory function tissue- or cell-type-specifically, into the analysis. We performed such analysis for the two arteries, as well as for data from several kidney cell types, using publicly available data from the ENCODE project and genome-wide genotypes from the Research Program on Genes, Environment and Health (RPGEH) Genetic Epidemiology Research on Adult Health and Aging (GERA) European-ancestry cohort. These analyses have revealed several genes

with strong expression support in the tissues of interest, and in multiple cases have also revealed tissue- or cell-type-specificity of these genes.

Chapter 2: Rare Coding Variants Associated with Blood Pressure Variation in 15,914 Individuals of African Ancestry

Introduction

Hypertension is a leading risk factor for cardiovascular disease. Several genome-wide association studies (GWAS) have implicated common variants at approximately 166 loci[12–35] associated with systolic (SBP) and diastolic (DBP) blood pressure (BP) across multiple studies and populations, but the specific genes involved are unknown and are yet to be narrowed down within each locus. Further, these common variants collectively explain only a small fraction (<3%) of the total phenotypic variance of BP traits.[39] There is an emerging consensus that low frequency and rare variants may account for a significant fraction of the remaining variance.[49] If so, genotyping catalogs different from those that enable GWAS are required. The Illumina HumanExome Beadchip ("exome chip") was designed as an array-based assay enriching for the lower frequency genetic variation that could be detected within exomes. The >240,000 variants on this chip were selected for their functional significance from ~12,000 human exome and genome sequences from individuals of varying ancestry (mainly European) and common disease states.[40] These variants include primarily non-synonymous, splice-site, and nonsense variants, with an enrichment of low frequency variants down to a frequency of occurrence at a minimum of two or three times in at least two datasets.[50]

The present study focuses on identifying specific genes associated with SBP and DBP long-term average (LTA) residuals, similar to our previous efforts with common variants,[18] adjusting for major BP confounders including body mass index (BMI), age,

and sex. We analyzed individuals of African ancestry within each study and then combined the results by meta-analysis across eight cohorts, including Africa America Diabetes Mellitus (AADM), Atherosclerosis Risk in Communities (ARIC) Study, Coronary Artery Risk Development in young Adults (CARDIA), Genetics Network (GenNet), Genetic Epidemiology Network of Arteriopathy (GENOA), Howard University Family Study (HUFS), Hypertension Genetic Epidemiology Network (HyperGEN), and Loyola University Chicago Cohort (LUC). The prevalence of hypertension is increased in African Americans as compared to other populations in the United States[51] and, moreover, they are not highly represented in genetic epidemiological studies[17,34,35,52]; the only previous major GWAS with an exclusively African-ancestry discovery sample size larger than ours focused on common variant-level analyses.[17,34] Therefore, this study aims to expand the list of genes potentially involved in BP outcomes in these individuals, with a focus on single variants across the full frequency spectrum, as well as on the gene level. These analyses suggest several rare variant gene candidates for BP regulation, which are likely candidates based on published functional data. Additional analyses focusing on a subset of genes known to be involved in monogenic hypertension and hypotension syndromes, in which many rare variants of large effect have been implicated in BP,[9,10,53] were also conducted.

Methods

Study Participants

Eight studies consisting of individuals (n=15,914 after quality control procedures) of African ancestry were analyzed in this study, including several individuals from the

family-based studies AADM,[54,55] GenNet,[56,57] GENOA,[56,57] HyperGEN,[56,57] and HUFS[58,59] as well as from the population-based cohorts ARIC,[60,61] LUC,[62–65] and CARDIA.[66,67] The AADM and LUC studies consist of native Africans with no European admixture; the participants from the remaining cohorts all consist of African Americans. The AADM cohort was designed to study the genetic epidemiology of diabetes and related traits including hypertension, and the HUFS cohort was designed to study multiple cardiometabolic traits including hypertension. The ARIC and CARDIA studies focus overall on cardiovascular traits and events, including blood pressure. The remaining studies were recruited specifically to study blood pressure and associated traits. Brief descriptions and additional data collection methods for each study are provided in the Supplemental Digital Content, Description of Cohorts. Genotyping information and phenotypic characteristics from the initial examination or phase of samples analyzed from each cohort are described in Table 1. An extended table describing phenotypic characteristics of all physical examinations/phases for all studies is presented as Supplemental Digital Content Table S1. All studies obtained written informed consent from the participants as well as approval from their institutional review boards.

Exome Chip Genotyping

Samples were genotyped either at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University on the Illumina HumanExome BeadChip v1.0 (247,870 variants), or at the Human Genetics Center at the University of Texas Health Science Center at Houston (Houston) on either v1.0 or v1.1 (242,901 variants) (http://www.chargeconsortium.com/main/exomechip).[68] Genotypes were called with

Illumina's GenTrain 1.0 (CIDR) or Gentrain 2.0 (Houston) clustering algorithm from the Illumina GenomeStudio v2011.1 software.

Quality control of Genotypes

Quality control (QC) of the samples and variants were carried out within each cohort, for this study, using the programs PLINK,[69] R[70] and KING.[71] We began with 17,940 total individuals and either 247,870 or 242,901 variants, depending on the version of the chip used. Technical failure variants were removed prior to sample and variant QC. Blind duplicates, identified by CIDR's QC reports where available, and unexpected duplicates, identified by PLINK's IBS distance metric, DST>0.98, were then removed from the samples, and those remaining were filtered on call rate (>95% retained), X-chromosome heterozygosity and PLINK's sex check F-statistic to identify sex discrepancies, as well as autosomal heterozygosity to identify individuals with unexpectedly high or low heterozygosity. Custom thresholds selected for the heterozygosity filters were a minimum of three standard deviations from the mean of each of the distributions. Finally, individuals with Mendelian errors, which may indicate genotyping errors or incorrect relationship assignment within their families, were excluded, as well as exhibiting too many first- or second- degree relationships within their cohort reflecting a data quality issue, as preliminary pedigree error correction; more extended pedigree error corrections were carried out after variant QC.

Quality control of variants include discarding those with >5% missing data, and those failing the Hardy-Weinberg Equilibrium (HWE) exact test in PLINK[3571] (defined as $p<1 \times 10^{-6}$ on variants with minor allele frequency (MAF)>0.01 only, as HWE estimates

on variants with lower MAF become less reliable).[34] Duplicate SNPs with discordant genotypes were removed within each cohort for downstream analyses.

Pedigree errors and sample switches were subsequently identified and correctd using KING and PLINK by identity-by-state (IBS) information. For the ARIC and LUC studies, only samples related as third degree or beyond were retained for analysis. These procedures left 15,914 individuals with ~233K variants in each cohort genotyped on the v1.0 chip and ~238K variants in each cohort genotyped on the v1.1 chip. The vast majority of the variants removed were due to technical failure. The average call rate of the samples remaining after all quality control procedures is ~99.95%.

Variant annotation

Variants on the chip were annotated using ANNOVAR[72] and phyloP[73] scores. A subset of 42,018 variants was annotated as "deleterious", defined as consensus intronic splice-site, nonsense-mediated decay (NMD)-compatible stop-gain, or conserved missense (phyloP > 4) variants. These variants were the ones considered in the Burden-T1-del analyses (refer to *Statistical Analyses* section).

Phenotypes

The traits SBP and DBP were analyzed in this study. For each sample, SBP and DBP measurements from all available visits were adjusted for medication use when applicable, by adding a fixed constant of 15 mmHg for SBP and 10 mmHg for DBP measurements.[74] These corrected measurements were then further adjusted in a linear

regression using the known blood pressure measurement confounders age, age-squared, sex, BMI, the first 10 principal components of ancestry, and study center (where applicable for each cohort) per available visit per individual as covariates. Principal components were calculated for all cohorts using KING-mds, which accounts for pedigree structure, on a linkage disequilibrium (LD)-pruned set (r^2<0.5 within window=50 SNPs, step=5 SNPs) of common variants (MAF>5%). The resulting adjusted BP values, or residuals, were then averaged across all available visits for each sample, and used as the phenotypes in the analyses described below. Selected individuals were dropped for issues including lower SBP than DBP, and implausible BMI values, but otherwise the full distribution was analyzed. The majority of individuals in some cohorts have data available for one examination only (AADM, GenNet, HUFS, HyperGEN, LUC), while in the others (ARIC, CARDIA, GENOA) the majority has data from more than one examination. The Pearson correlation coefficients for first examination residual versus average residual for individuals with multiple visits in each cohort are reported in Supplemental Digital Content Table S2.

Statistical Analyses

Study-level analyses were conducted, followed by fixed-effects meta-analyses to combine the results with the seqMeta[75] package in R, using SBP and DBP residuals as phenotypes. To account for correlations among relatives, a kinship matrix was used in seqMeta for family-based cohorts. For these meta-analyses, the union of non-monomorphic SNPs passing QC in each of the cohorts was analyzed (170,540 variants). Analyses were restricted to autosomal and X-chromosomal variants.

We performed both single-variant analyses assuming an additive model, as well as gene-based tests, using the included SNP information files in seqMeta for gene definitions. The gene-based tests are particularly useful for analysis of rare variants, which individually have low numbers of carriers in the population, by aggregating variants into functional units (genes) and testing for cumulative effects. The single variant analyses were grouped into three frequency classes (common: MAF≥0.05, low frequency: 0.01≤MAF<0.05 and rare: MAF<0.01). For gene-based analyses, we considered sequence-kernel association test (SKAT),[43] the T1 burden test[42] including all variants with MAF<0.01 (Burden-T1-all), and the T1 burden test including only those variants predicted to be deleterious (Burden-T1-del), which are defined as intronic splice, NMD-compatible stop-gain, or conserved missense (phyloP>4). The SKAT test is expected to be more powerful when the region contains variants that are largely neutral or have opposing directional effects, while burden tests are expected to be more powerful when the region contains variants largely in the same direction.[43,76] Statistical significance was defined by the Bonferroni method of multiple test correction, considering the number of associations computed within each analysis. There was no additional correction for number of phenotypes (two) due to the high correlation between the phenotypes (Supplemental Digital Content Table S3: Pearson's r, 0.75-0.87).

Power Calculations

Assuming an additive model, power was calculated as described previously.[38] Under an additive model, a quantitative trait can be represented by a two-component normal mixture distribution. One component represents the reference allele, weighted by

its frequency, p, and the other the alternate allele, with a genetic effect shifted by s standard deviations with respect to the reference allele distribution, weighted by its frequency q (q=1-p). Both distributions are assumed to have variance of 1. Where n is sample size, Φ is the standard normal cumulative distribution function (CDF), and $z_{1-\alpha/2}$ is the standard normal distribution quantile at significance level α, the power to detect the difference in means of the two distributions is calculated as:

$$\Phi\left(\frac{s}{\sqrt{\left(\frac{1}{2n(1-q)}+\frac{1}{2nq}\right)}}-z_{1-\frac{\alpha}{2}}\right)+\Phi\left(\frac{-s}{\sqrt{\left(\frac{1}{2n(1-q)}+\frac{1}{2nq}\right)}}-z_{1-\frac{\alpha}{2}}\right)$$

Results

We conducted single variant analyses by frequency class (common, low frequency and rare), as well as gene-based SKAT and T1 burden tests as defined above on 170,540 variants in 15,914 individuals of African ancestry. Singleton variants (n=18,217) were dropped in the single rare-variant analyses to prevent observing significance resulting from partiality due to phenotypic outliers. Further, in gene-based tests, genes with only a single variant (n=10,041) were excluded, as they were already included in single-variant analysis, allowing assessment of only those that may possibly reflect an aggregate effect of variants. Genomic inflation factors were calculated for all 170,540 single variants analyzed as well as the gene-based test results, for each of the cohorts and for the meta-analyses. The inflation factors for all of the meta-analyses are

shown in Supplemental Digital Content Table , and quantile-quantile (QQ) plots for meta-analyses are displayed in Supplemental Digital Content Figures S1-S2. The inflation factors for these analyses for each cohort are provided in Supplemental Digital Content Table S5. The LUC cohort exhibits slight inflation for a number of the tests; however, no additional adjustment was made, as all inflation factors fall near or within the acceptable range for GWAS.[77,78] Otherwise, minimal inflation was observed across most cohorts, even at the meta-analysis level, although the QQ plots for rare variants presents several variants deviated from the expected, despite well-controlled genomic inflation factors.[79] The inflation factor uses the median, which is robust to outliers and so may not be aligned with the QQ plot when a specific set of outliers is present.

A brief summary of the number of tests conducted, the Bonferroni correction threshold, and significant results for each analysis are shown in Table 2. A total of 152,323 (non-singleton) variants were used in the single variant analyses. Significant results from the single variant results are outlined in Table 3, and those from the gene-based tests are outlined in Table 4. Although the seqMeta program does not use approximated effect sizes (betas) in its score test, it provides estimates of the (non-standardized) betas and their standard errors. The beta estimates for all of these significant rare variants indicate a BP *increasing* effect.

Single-variant analyses of 105,487 variants in the rare variant class revealed several associations with SBP and DBP (Bonferroni threshold $p=4.74 \times 10^{-7}$), summarized in Table 3, with supporting power calculations in Supplemental Digital Content Table S6, and depicted in the Manhattan plots in Figure 1. The vast majority of these variants are non-synonymous (missense). Significant associations of rs150432347

in *COL6A1* ($p=1.19 \times 10^{-7}$), with 19 copies in four cohorts, as well as rs138594727 in *CRYBA2* ($p=3.02 \times 10^{-7}$), with 15 copies in six cohorts, were identified with SBP. The intronic splice-site SNP rs11568416 in the gene *SLC28A3* is additionally associated with SBP ($p=8.92 \times 10^{-8}$), with six copies present across unrelated individuals in four cohorts. Other single-variant analysis results include similar associations of rare variants in *KRBA1, SEL1L3,* and *YOD1* with SBP, each with two to four carriers. The variant rs148474705 in *GAPDHS* is associated with both SBP and DBP, also with two carriers, and the SNP rs142319329 in the gene *AFF1* is associated with DBP ($p=2.95 \times 10^{-7}$), with eight copies present across unrelated individuals in four cohorts. These variants are all present in heterozygous carriers. The mean standardized effect sizes generally indicate that more copies of the variant correlate with smaller effects, but results in this study are too sparse for this conclusion to be well supported.

Table 4 contains the results for the two significant genes from the gene-based tests (Bonferroni threshold $p=1.02 \times 10^{-5}$), which are depicted in the Manhattan plots in Figure 2. The T1 test on deleterious variants identified significant associations of *CCDC13* for SBP ($p=3.54 \times 10^{-7}$) and *QSOX1* for DBP ($p=3.86 \times 10^{-6}$). The associations of *CCDC13* with DBP (6.90×10^{-5}) and *QSOX1* with SBP ($p=3.09 \times 10^{-5}$) are also close to statistical significance. This analysis included two SNPs for *CCDC13*, each with one copy in different cohorts: rs182436192 (SBP single variant $p=5.66 \times 10^{-3}$) and rs143310118 (SBP single variant $p=1.90 \times 10^{-5}$). There were also two SNPs included in the analysis for *QSOX1*: rs202144688, with one copy (DBP single variant $p=7.59 \times 10^{-4}$), and rs201390473, with two copies (DBP single variant $p=9.26 \times 10^{-4}$) in different cohorts.

The low frequency, the common single variant analyses, SKAT and Burden-T1-all analyses produced no significant results for either phenotype. Manhattan plots for these analyses are shown in Supplemental Digital Content Figures S3 and S4. However, it may be noted that the top result from the low frequency analysis for DBP was rs73828047 from *ULK4* ($p = 1.79 \times 10^{-5}$); further, the *ULK4* gene is second on the list of top results for the SKAT analysis in DBP ($p=4.54 \times 10^{-5}$), with rs73828047, rs2272007 (2.64×10^{-3}), rs1052501 ($p=2.75 \times 10^{-3}$), rs1716975 ($p=3.34 \times 10^{-3}$), and rs192994614 ($p=6.34 \times 10^{-3}$) as the top five SNPs. These results are particularly of interest, as rs2272007 (MAF=0.30) and rs1716975 (MAF=0.30) are among the common variants previously associated (or in high LD with those associated) with DBP in multiple ethnicities,[17,26] including African Americans. We additionally examined a set of 212 variants representative of the 166 BP GWAS loci, of which 80 were represented in our final cleaned set. While our SBP single variant results fail to support any of these 80 variants, a few show evidence at $p<0.01$ in the DBP single variant results: rs6969780 in *HOXA3* ($p=5.37 \times 10^{-4}$, MAF=0.32), rs6825911 in *ENPEP* ($p=1.70 \times 10^{-3}$, MAF=0.45), rs1925153 in *COL21A1* ($p=4.51 \times 10^{-3}$, MAF=0.41), and rs926552 in *SNORD32B* (8.79×10^{-3}, MAF=0.13). These results, though not significant in these analyses, may lend additional credence to the possibility of roles of these loci in BP regulation.

Assessment of gene-based test results for the known monogenic renal salt-handling BP genes surprisingly provided little evidence of their roles in BP regulation. The SKAT test for DBP produced the lowest p-value for the potassium channel gene, *KCNJ5* ($p=1.12 \times 10^{-3}$). Four SNPs were analyzed in the SKAT test, but single-variant p-values suggest that this evidence is primarily driven by the rs115012103 SNP ($p=9.95 \times$

10^{-4}). The SBP SKAT test p-value ($p=4.65 \times 10^{-3}$) for this gene is similarly evidential. Results from variant-based tests for this gene are listed in Supplemental Digital Content Table S7. Overall, the lack of support for the monogenic genes may possibly be attributed to their explaining a small proportion of all essential hypertension cases.

Discussion

This meta-analyses of eight cohorts with individuals of African ancestry have identified several statistically significant associations of rare variants with SBP and DBP (*AFF1*, *GAPDHS*, *SLC28A3*, *COL6A1*, *CRYBA2*, *KRBA1*, *SEL1L3*, *YOD1*, *CCDC13*, *QSOX1*), and, further, provided additional evidence for previously identified BP loci. Beyond these statistical results, prior biological experiments offer compelling support for the plausibility of these candidates for BP regulation, presented in Table 5.[80–92]

There were no significant results from either the common or low frequency single variant analyses, or from the SKAT and Burden-T1-all analyses. Given that the focus of the array is on rare, functional variants, the significant results are, as expected, from the rare single variant and Burden-T1-del analyses. Additionally, the majority of the significant rare variants have just two to three copies across the eight cohorts analyzed in this study; this is also expected as one criterion for variant selection, as stated before, is that the majority of these variants only be present in two to three copies in a minimum of two studies, across multiple populations. Supplemental Digital Content Table S8 details frequencies of these alleles in the African ancestry sample from the Exome Aggregation Consortium (ExAC), with similar numbers of copies and frequencies as in this study. Further, the consistency of measurements across multiple visits in a study

where applicable for carriers augments the verity of the associations. Additionally, the power calculations in Table S6 support that there is sufficient statistical power (~55-80%+) to detect the majority of variants of these effects and frequencies in this sample size. Despite this, it still must be considered that the few numbers of copies, especially those with only two copies, of these significant results strongly indicate that they require replication and further study for additional support.

Though there is some evidence for the ~166 GWAS replicated loci as well as the Mendelian syndromic blood pressure genes, their lack of significance in this study is surprising. It is likely that the dearth of common variants available on the chip contributed to inadequate tagging of causal variants within these loci. Despite this, it should be noted that recent BP exome array studies by Liu et al.[31] and Surendran et al.[32] replicated several common, as well as low-frequency, variants. However, these studies consisted of mostly European ancestry participants, and as the sources of sequences that this chip was designed from were primarily European individuals, this likely made the array less informative for individuals of African and other ancestries. An ethnicity-specific chip may therefore be more revelatory, as the familial hypertension variants might be different in African ancestry patients. Additionally, these studies consisted of over 300,000 individuals combined across the discovery and replication stages, while our study consisted of just ~16,000 individuals, at which power is much lower to detect common variants at their typical effect sizes.

The Mendelian genes are contrastingly well known for containing rare causal variants with large effects, though we saw poor signal here as well. As stated above, an ethnicity-specific array may be more helpful here as well. Another possibility to explain

the absence of a signal in these genes is that their effect size might be somewhat smaller than the one observed in the rare variants of this experiment, and therefore the statistical experimental power is insufficient to show an association signal.

Regardless, it is of interest that rare variants identified in this study may contribute to blood pressure variation as a polygenic trait. Studies of schizophrenia have identified rare variants from exome-based studies,[93,94] demonstrating an enrichment of rare variants in genes containing common variants implicated in previous schizophrenia GWAS, aiding in fine-mapping of those loci. In our study, though the results did not reach statistical significance, the statistical support for low-frequency variants in *ULK4* provides evidence for this as the specific gene within its SBP/DBP GWAS locus.

In summary, we identified several rare variants in 10 genes (*AFF1*, *GAPDHS*, *SLC28A3*, *COL6A1*, *CRYBA2*, *KRBA1*, *SEL1L3*, *YOD1*, *CCDC13*, and *QSOX1*) that are significantly associated with SBP and DBP traits in 15,914 individuals of African ancestry. In contrast to the previously identified common variants of most BP GWAS, and the rare variants in monogenic blood pressure syndromes, this study has identified rare variants that are potentially contributory to blood pressure as a polygenic trait, particularly in individuals of African-ancestry. The prior experimental evidence of the involvement of these genes with related traits in animal models suggests that these genes are viable candidates for BP regulation, and will benefit from additional investigation. Additionally, the results of this study suggest that future studies relying on genotype arrays for analysis would offer more substantial findings using ethnicity-specific chips.

Acknowledgements

We thank the participants of all of the studies involved and assistance of the research staff and all investigators that have made our analyses possible.

Chapter 2 Figures

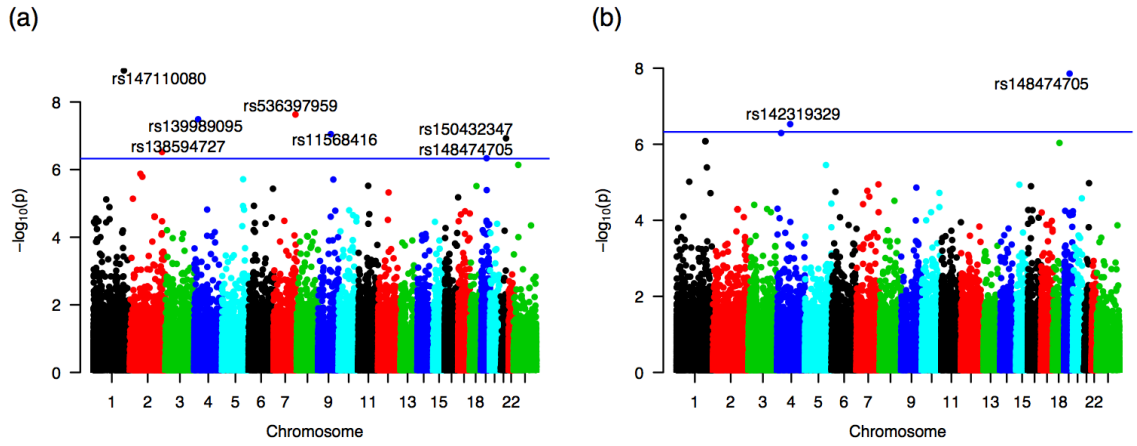

Figure 1. Manhattan Plots for 105,487 rare single variant analyses of SBP (a) and DBP (b).[95]

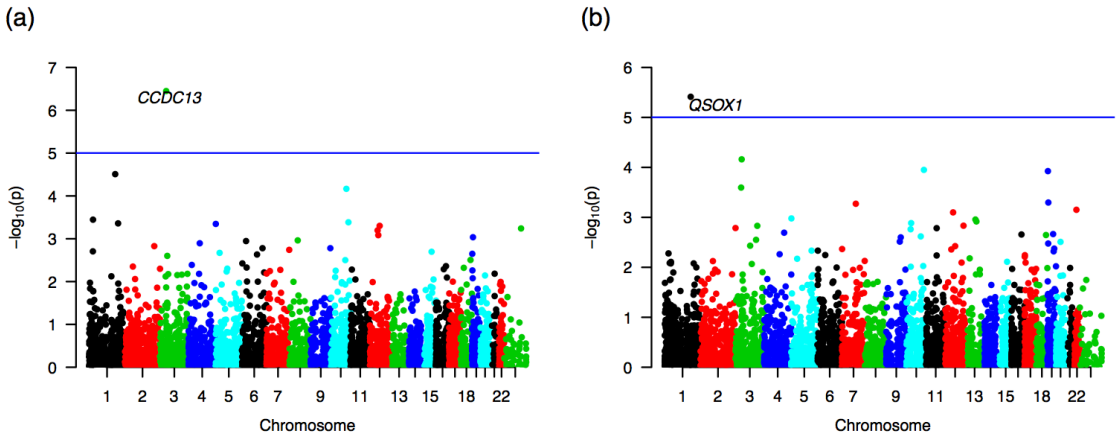

Figure 2. Manhattan Plots for T1 analyses of deleterious variants at 5,024 genes for SBP (a) and DBP (b). Each gene is plotted at the position of its median SNP available on the chip.

Chapter 2 Tables

Table 1. Genotypic and phenotypic characteristics for initial examination[a] in the eight cohorts analyzed.

Cohort	Chip Version	# analyzed	% Female	Mean Age (SD)	Mean BMI (SD)	Mean SBP (SD)	Mean DBP (SD)	% on anti-hypertensive medication
AADM	v1.1	2070	57	45.55 (16.02)	25.96 (5.821)	132.69 (22.65)	81.01 (14.19)	28
ARIC	v1.0/v1.1	3095	63	53.61 (5.80)	29.72 (6.26)	128.63 (21.00)	79.65 (12.15)	41
CARDIA	v1.0	1534	58	31.45 (3.80)	28.24 (6.82)	111.33 (13.01)	70.92 (10.97)	3.2
GenNet	v1.1	782	60	40.42 (11.45)	30.25 (8.12)	127.09 (20.20)	77.70 (13.60)	100
GENOA	v1.1	1373	69	56.67 (10.66)	31.02 (6.57)	131.69 (23.51)	70.96 (11.53)	100
HUFS	v1.1	1784	61	43.41 (14.44)	30.69 (8.672)	127.38 (20.99)	80.18 (12.87)	27
HyperGEN	v1.1	1798	65	48.56 (12.84)	32.09 (7.61)	130.10 (22.31)	74.67 (12.00)	100
LUC	v1.1	2688	71	48.66 (11.72)	26.12 (6.17)	146.09 (29.98)	91.92 (18.64)	39

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; SD, standard deviation
[a]As this only illustrates the first visit or phase, there are many individuals in subsequent examinations that are included in these analyses but not in this table; the complete information may be found in Table S1.

Table 2. Summary of number of significant results and tests per analysis.

Test	# tests	Bonferroni Corrected Threshold[a]	# passing threshold (SBP)	# passing threshold (DBP)
Common	28,851	1.73×10^{-6}	0	0
Low frequency	17,985	2.78×10^{-6}	0	0
Rare	105,487	4.74×10^{-7}	7	2
Gene-based	15,554	3.21×10^{-6}	0	0
T1 (all variants)	14,465	3.46×10^{-6}	0	0
T1 (deleterious variants)	5,024	9.95×10^{-6}	1	1

SBP, systolic blood pressure; DBP, diastolic blood pressure; # tests, number of variants or genes analyzed

[a]The significance threshold as determined by the Bonferroni method corrects for the number of tests (0.05/# tests) in each analysis.

Table 3. Rare Single Variant Hits for SBP and DBP ($p<4.74 \times 10^{-7}$).

Variant	Gene	Site	Type	Trait	MAF	p-value	# copies	β_{sm} (SE)[a]	β_{st} (95%CI)[b]
rs11568416	*SLC28A3*	intronic	splice	SBP	1.89×10^{-4}	8.92×10^{-8}	6	39.44 (7.38)	1.96 (1.16 – 2.76)
rs536397959	*KRBA1*	exonic	synonymous	SBP	6.28×10^{-5}	2.35×10^{-8}	2	73.65 (13.19)	3.93 (2.55 – 5.32)
rs139989095	*SEL1L3*	exonic	missense	SBP	6.28×10^{-5}	3.23×10^{-8}	2	49.02 (8.87)	3.91 (2.53 – 5.30)
rs147110080	*YOD1*	exonic	missense	SBP	6.28×10^{-5}	1.18×10^{-9}	2	64.41 (10.59)	3.57 (2.18 – 4.96)
rs150432347	*COL6A1*	exonic	missense	SBP	6.05×10^{-4}	1.19×10^{-7}	19	22.14 (4.18)	1.16 (0.71 – 1.61)
rs138594727	*CRYBA2*	exonic	missense	SBP	4.73×10^{-4}	3.02×10^{-7}	15	25.92 (5.06)	1.27 (0.76 – 1.78)
rs148474705	*GAPDHS*	exonic	missense	SBP	6.28×10^{-5}	4.60×10^{-7}	2	53.85 (10.68)	3.51 (2.12 – 4.89)
rs148474705	*GAPDHS*	exonic	missense	DBP	6.28×10^{-5}	1.39×10^{-8}	2	40.24 (7.09)	3.97 (2.58-5.36)
rs142319329	*AFF1*	exonic	missense	DBP	2.51×10^{-4}	2.95×10^{-7}	8	21.10 (4.12)	1.77 (1.07 – 2.46)

SBP, systolic blood pressure; DBP, diastolic blood pressure; MAF, minor allele frequency; β_{sm}, seqMeta beta; β_{st}, beta as standardized mean difference; SE, standard error; 95% CI, 95% confidence interval; # copies, allele count

[a]The β_{sm} (SE) values are reported from seqMeta output, in units of the phenotype residuals.

[b]The β_{st} (95% CI) values are calculated as the mean standardized difference between the LTA residuals for all carriers and all non-carriers for each variant.

Table 4. Burden-T1-del significant genes for SBP and DBP ($p<9.95 \times 10^{-6}$).

Gene	Test	Trait	p-value	# SNPs analyzed	Cumulative Carrier Count	β_{sm} (SE)[a]
CCDC13	Burden-T1-del	SBP	3.54×10^{-7}	2	2	54.38 (10.68)
QSOX1	Burden-T1-del	DBP	3.86×10^{-6}	2	3	32.93 (7.13)

Burden-T1-del, T1 burden test on deleterious variants with minor allele frequency (MAF)<0.001; # SNPs analyzed, number of SNPs included in test for gene; Cumulative Carrier Count, number of carriers (same as allele count) across all analyzed variants; β_{sm}, seqMeta beta; SE, standard error

[a]The β_{sm} (SE) values are reported from seqMeta output, in units of the phenotype residuals.

Table 5. Evidence of genes containing sentinel variants for roles in hypertension.

Gene	Evidence of role in hypertension
COL6A1	Encodes a collagen VI protein and provides structural support for a variety of tissues including the heart; has been shown previously, in addition to other VEGF-A pathway genes, to be associated with atrioventricular septal defect (AVSD) in patients with Down syndrome[80]
SLC28A3	Encodes a sodium-dependent nucleoside transporter (NT); NTs have many physiological regulatory roles including that of mediating adenosine concentration, which in turn affects vascular tone[81,82]
SEL1L3	Differentially expressed between male familial combined hyperlipidemia and coronary heart disease (FCHL-CHD) patients and non-FCHL-non-CHD controls in a microarray study.[83] Also upregulated with a >6-fold expression change in a rabbit microarray study comparing simulating conditions before and after repair of coarctation of the aorta, a condition which may evolve into chronic hypertension[84]; as the authors stated, despite the surgically induced return to normal BP, the vasculature still retains its physical defects, and this gene may thus play a role in the continued residual effects.
YOD1	A deubiquitinase, targeted by miR-21 in the distal small pulmonary arteries of mice with pulmonary arterial hypertension exposed to hypoxic conditions, as well as in human pulmonary artery smooth muscle cells with transfected miR-21.[85]
CCDC13	Encodes a centriolar satellite protein involved in primary ciliogenesis[64]; this supports a potential role in hypertension present in individuals with autosomal dominant polycystic kidney disease, in which impaired primary cilia affects vascular tone[87,88]
QSOX1	Recent experiments support cardiovascular function for QSOX1, encoding a sulfhydryl oxidase enzyme; one such function includes its induction of vascular smooth muscle cell migration and proliferation in vitro,[89] a potential role in atherosclerosis,[90] and its feasibility as a biomarker for acute decompensated heart failure.[91,92]

Chapter 2 Supplement

Supplemental Digital Content. Description of Cohorts; Supplementary Tables S1-S8 describing population demographics, phenotype correlations (LTA and SBP with DBP), genomic inflation factors for statistical tests, statistical power for the tests, *KCNJ5* single variant test results, and ExAC frequencies for rare variants; Supplementary Figures S1-S4 depicting QQ and Manhattan plots for non-significant results in this study; Complete Acknowledgements.

Descriptions of Cohorts

GenNet, GENOA and HyperGEN (FBPP)

The GenNet, GENOA (Genetic Epidemiology Network of Arteriopathy) and HyperGEN (Hypertension Genetic Epidemiology Network) cohorts are family-based studies that are part of the Family Blood Pressure Program (FBPP),[56,57] with visit dates from 1995 to 2003. For these studies, hypertension was defined as SBP ≥140 mmHg and/or DBP ≥90 mmHg, and/or diagnosis and treatment with prescribed hypertension medication; severe hypertension was defined as SBP ≥160 mmHg and/or DBP ≥100 mmHg, or taking two or more hypertensive medications.

The GenNet study consists of 1,101 African-Americans and 839 Hispanic-Americans from Maywood, IL, as well as 1,497 European-Americans from Tecumseh, MI, studied over two phases. This study has exome chip data available from 962 African-Americans from 338 families in the final dataset. Probands aged 18-50 years with measured blood pressure values in the top 20-25th percentiles, as well as their siblings and parents wherever possible, were recruited into the study.

The GENOA study primarily focuses on siblings diagnosed with hypertension prior to age 60. The study includes data on 1,606 European-Americans from Rochester, MN, 1,857 African-Americans from Jackson, MS, and 1,803 Hispanic-Americans from Starr County, TX, studied across two phases. In this analysis, exome chip data from 1,373 African-Americans, comprising 562 families, were used for analysis.

The HyperGEN study includes 2,471 European-Americans and 2,300 African-Americans from Birmingham, AL, Forsyth County, NC, Framingham, MA, Minneapolis, MN, and Salt Lake City, UT. The cohort consists preferentially of sibships containing at least two individuals with hypertension of which one individual had severe hypertension, as well as their parents wherever possible, and a minimum of one untreated adult offspring. There are also ~800 random subjects age-matched with individuals in hypertensive sibships. Exome chip data from 2,059 African-Americans, comprising 902 families, are available in the final data for analysis.

For both SBP and DBP, measurements were taken on an oscillometric automated BP device. Either three measurements were taken on a Dinamap instrument, or two measurements were taken on an Omron instrument. In GenNet individuals, measurements were taken on either a Dinamap or an Omron instrument; in GENOA and HyperGEN, all measurements were taken on a Dinamap instrument. An average of up to three (Dinamap) or two (Omron) measurements of SBP and DBP were used in this analysis.

LUC

Participants of the LUC (Loyola University Chicago) cohort were recruited from Yoruba-speaking communities in southwest Nigeria, specifically Igbo-Ora and Ibadan, as part of a long-term study on the environmental and genetic factors underlying hypertension.[62,63] The study protocol was reviewed and approved by the Institutional Review Board at Loyola University Medical Center, and the Joint Ethical Committee of the University of

Ibadan/University College Hospital, Ibadan, Nigeria. All participants gave written informed consent administered in either English or Yoruba. Phenotype measurements were performed by trained research staff using a standardized protocol.[64,65] Body weight was measured to the nearest 0.2 kg on calibrated electronic scales, whereas height was obtained using a stadiometer consisting of a steel tape attached to a straight wall and a wooden headboard. Systolic and diastolic blood pressures were measured using an Omron oscillometric device. Three measurements were taken three minutes apart and the average of the final two was used in the analysis. Participants with hypertension were offered treatment after detection at the screening examination. A total of 2,688 unrelated adults with quality controlled exome chip data were included in the present study.

HUFS

The HUFS (Howard University Family Study) is a multi-generational family-based study of African-American individuals enrolled from the Washington, DC metropolitan area.[58,59] SBP and DBP were measured using an Omron oscillometric BP device. Three measurements were taken in the sitting position, with ten minutes between each measurement, with the average of the last two measurements being used in this analysis. There are exome chip data from 1,784 individuals in 1,139 families available for analysis.

AADM

The AADM (Africa America Diabetes Mellitus) study enrolled both families and unrelated West African individuals focusing on Type 2 Diabetes from five study centers, including Accra and Kumasi in Ghana, and Enugu, Ibadan and Lagos in Nigeria. The first

and second phases of the study recruited affected sibling pairs and spouse controls and the third phase enrolled unrelated cases and controls. Information on several diabetes and cardiovascular-related traits[54] were measure during clinic visit. Blood pressure was measured with 10 minutes between each measurement, with the average of the last two measurements being used in this analysis, as described for HUFS.[55] There are 2,070 individuals from 1,202 families present in the final dataset.

ARIC

The ARIC (Atherosclerosis Risk in Communities) study is a population-based, prospective study on 15,792 individuals, including 11,478 European-Americans and 4,266 African-Americans from Forsyth County, NC, Jackson, MI, Minneapolis, MN and Washington County, MD.[60,61] There have been five visits, with individuals in the first visit, in 1987-1989, being of ages between 45 and 64 years. Subsequent visits occurred in 1990-1992, 1993-1995, 1996-1998, and 2011-2013. The data from the first four visits were used in this analysis. SBP and DBP were measured thrice at each of the first three visits, and twice at the fourth visit, using a random zero sphygmomanometer; the average of the (final) two measurements were used for analysis. In this study, exome chip data from 3,280 African-Americans are available in the final cleaned dataset.

CARDIA

The CARDIA (Coronary Artery Risk Development in young Adults) study is a population-based, prospective study of 5,115 young European-Americans and African-Americans, with data collected on traits related to cardiovascular disease.[66,67] These

individuals were selected from four study centers: Birmingham, AL, Chicago, IL, Minneapolis, MN, and Oakland, CA. The study began with the first examination in 1985-1986 during which participants were 18-30 years old. Seven subsequent examinations were conducted in years 2, 5, 7, 10, 15, 20, and 25, with the most recent examination in 2010-2011. The data from years 7, 10, 15 and 20 were used in this analysis, and additionally only individuals with at least two visits of data were included in this analysis. There are 1,698 African-American individuals analyzed in the final cleaned exome chip dataset.

Table S1. Phenotypic characteristics for all examinations per cohort.

Cohort	Visit/ Phase	# individuals	% Female	Mean Age (SD)	Mean BMI (SD)	Mean SBP (SD)	Mean DBP (SD)	% on anti-hypertensive medication
AADM	1	2070	57	45.55 (16.03)	25.96 (5.82)	132.69 (22.65)	81.01 (14.19)	28
ARIC	1	3095	62	53.61 (5.80)	29.72 (6.26)	128.63 (21.00)	79.65 (12.15)	41
ARIC	2	2706	63	56.41 (5.79)	30.05 (6.35)	127.17 (21.07)	75.45 (11.29)	45
ARIC	3	2274	63	59.11 (5.70)	30.44 (6.50)	131.38 (21.16)	76.34 (11.05)	49
ARIC	4	1989	63	61.88 (5.68)	30.67 (6.49)	134.15 (20.60)	75.94 (10.76)	55
CARDIA	1	1534	58	31.45 (3.80)	28.24 (6.82)	111.33 (13.01)	70.92 (10.97)	3.2
CARDIA	2	1567	58	34.39 (3.83)	29.11 (7.10)	112.96 (13.86)	74.76 (10.72)	5.8
CARDIA	3	1439	59	39.56 (3.84)	30.56 (7.47)	117.38 (16.29)	77.19 (12.72)	13.1
CARDIA	4	1341	61	44.54 (3.84)	31.43 (7.69)	120.78 (16.32)	76.68 (11.71)	25.7
GenNet	1	782	60	40.42 (11.45)	30.25 (8.12)	127.09 (20.20)	77.70 (13.60)	100
GenNet	2	182	60	32.13 (13.11)	29.86 (8.54)	116.76 (16.72)	67.58 (13.76)	13
GENOA	1	1373	69	56.67 (10.66)	31.02 (6.57)	131.69 (23.51)	70.96 (11.53)	100
GENOA	2	1082	71	61.45 (9.80)	31.79 (6.70)	134.10 (22.31)	73.35 (10.57)	67

HUFS	1	1784	61	43.41 (14.44)	30.69 (8.672)	127.38 (20.99)	80.18 (12.87)	27
HyperGEN	1	1798	65	48.56 (12.84)	32.09 (7.61)	130.10 (22.31)	74.67 (11.99)	100
HyperGEN	2	261	60	32.81 (8.29)	31.69 (8.56)	119.02 (17.45)	71.26 (10.75)	6.1
LUC	1	2688	71	48.66 (11.72)	26.12 (6.17)	146.09 (29.98)	91.92 (18.64)	39

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; SD, standard deviation

Table S2. Pearson correlations of first available visit with LTA (long-term average) residuals per cohort[a]

Cohort	# with multiple visits	SBP residual correlation	DBP residual correlation
AADM	0	NA	NA
ARIC	2806	0.85	0.85
CARDIA	1698	0.73	0.70
GenNet	2	NA	NA
GENOA	1079	0.86	0.87
HUFS	0	NA	NA
HyperGEN	0	NA	NA
LUC	0	NA	NA

SBP, systolic blood pressure; DBP, diastolic blood pressure

[a]Pearson correlations between the first visit and LTA residuals for all individuals with data for multiple visits (# with multiple visits) are shown here. This comparison is largely meaningful for GENOA, ARIC, and CARDIA, for which for the majority of individuals have available data for more than one visit.

Table S3. Pearson correlations between SBP and DBP residuals.

Cohort[a]	Correlation
AADM	0.77
ARIC	0.76
CARDIA	0.81
GenNet	0.77
GENOA	0.75
HUFS	0.79
HyperGEN	0.76
LUC	0.87

[a]Either one visit or LTA where available: (GENOA, ARIC, and CARDIA) residuals are shown here per cohort.

Table S4. Genomic control inflation factors (λ_{GC}) for meta-analyses.

Test	λ_{GC} (SBP)	λ_{GC} (DBP)
Single Variant[a]	1.036	1.026
SKAT	1.082	1.053
Burden-T1-all	1.011	1.010
Burden-T1-del	0.986	0.982

λ_{GC}, genomic inflation factor; SBP, systolic blood pressure; DBP, diastolic blood pressure

[a]The overall inflation factor is shown here for all 170,540 single variants. Inflation factors by frequency class are provided in the QQ plots in Figure S1.

Table S5. λ_{GC} for all analyses per cohort.

Cohort	SBP SS	SBP SKAT	SBP Burden-T1-all	SBP Burden-T1-del	DBP SS	DBP SKAT	DBP Burden-T1-all	DBP Burden-T1-del
AADM	0.997	0.829	1.017	0.996	0.981	0.789	0.993	1.017
ARIC	0.994	0.952	1.015	0.985	1.021	0.970	0.998	0.997
CARDIA	0.989	0.888	0.941	0.901	0.986	0.874	0.985	0.963
GenNet	0.911	0.747	0.960	0.883	0.945	0.777	0.971	0.899
GENOA	0.997	0.853	1.041	0.929	0.965	0.838	1.007	0.940
HUFS	0.966	0.865	0.985	0.969	0.953	0.824	0.966	0.956
HyperGEN	0.989	0.875	1.009	1.001	0.960	0.858	0.994	0.954
LUC	1.029	0.833	1.043	1.048	1.064	0.877	1.074	1.061

SBP, systolic blood pressure; DBP, diastolic blood pressure; SS, single variant test; Burden-T1-del, T1 burden test on all variants with minor allele frequency (MAF)<0.01; Burden-T1-del, T1 burden test on deleterious variants with MAF<0.01

Table S6. Power[a] (%) for rare variants with mean standardized effect sizes 2, 3, and 4 in 15,914 individuals

AF[b]/ES[c]	Num copies[b]	2	3	4
3.141×10^{-5}	1	0.12	2.085	14.996
6.283×10^{-5}	2	1.361	21.358	73.241
9.430×10^{-5}	3	5.79	56.331	97.071
1.257×10^{-4}	4	14.991	83.224	99.848
1.517×10^{-4}	5	28.612	95.265	99.995

AF, allele frequency; ES, effect size

[a]Power (%) is shown in columns 3-5 as a function of allele frequency and effect size

[b]Allele frequency is presented in each row (column 1), with the corresponding allele count (Num copies) the frequency represents (column 2)

[c]Effect size is presented in each column (row 1), presented in standard deviation units (effects of 2, 3, and 4 standard deviations)

Table S7. *KCNJ5* Single Variant results for SBP and DBP.

rsID_dbSNP137	Trait	p-value	MAF	β_{sm} (SE)[a]
rs115012103	SBP	3.98×10^{-3}	0.011	3.166 (1.099)
rs115012103	DBP	9.95×10^{-4}	0.011	2.220 (0.674)
rs139073333	SBP	0.844	0.001	0.826 (4.199)
rs139073333	DBP	0.741	0.001	0.872 (2.640)
rs138295501	SBP	0.809	0.000	4.887 (20.208)
rs138295501	DBP	0.992	0.000	0.135 (12.948)
rs148307402	SBP	0.482	0.001	2.997 (4.263)
rs148307402	DBP	0.530	0.001	1.506 (2.397)

SBP, systolic blood pressure; DBP, diastolic blood pressure; rsID_dbSNP137, rs number variant ID from dbSNP137; MAF, minor allele frequency; β_{sm}, seqMeta beta

[a]The β_{sm} (SE) values are reported from seqMeta output, in units of the phenotype residuals.

Table S8. Allele frequencies for rare significant SNPs from the ExAC database on African Population.

Gene	Variant	Trait	Meta-analysis MAF	# copies	ExAC African MAF[a]	ExAC # copies
SLC28A3	rs11568416	SBP	1.89×10^{-4}	6	3.91×10^{-4}	4
KRBA1	rs536397959	SBP	6.28×10^{-5}	2	NA	NA
SEL1L3	rs139989095	SBP	6.28×10^{-5}	2	1.02×10^{-4}	1
YOD1	rs147110080	SBP	6.28×10^{-5}	2	9.64×10^{-5}	1
COL6A1	rs150432347	SBP	6.05×10^{-4}	19	6.58×10^{-4}	6
CRYBA2	rs138594727	SBP	4.73×10^{-4}	15	6.48×10^{-4}	6
GAPDHS	rs148474705	SBP/DBP	6.28×10^{-5}	2	1.95×10^{-4}	2
AFF1	rs142319329	DBP	2.51×10^{-4}	8	3.20×10^{-4}	3
CCDC13	rs182436192	SBP	3.14×10^{-5}	1	NA	NA
	rs143310118		3.14×10^{-5}	1	NA	NA
QSOX1	rs202144688	DBP	3.14×10^{-5}	1	1.01×10^{-4}	1
	rs201390473		6.28×10^{-5}	2	NA	NA

SBP, systolic blood pressure; DBP, diastolic blood pressure; MAF, minor allele frequency; # copies, allele count in population

[a]Though a few of these variants are not present in the African ExAC sample, they may exist in other ExAC samples (not noted here).

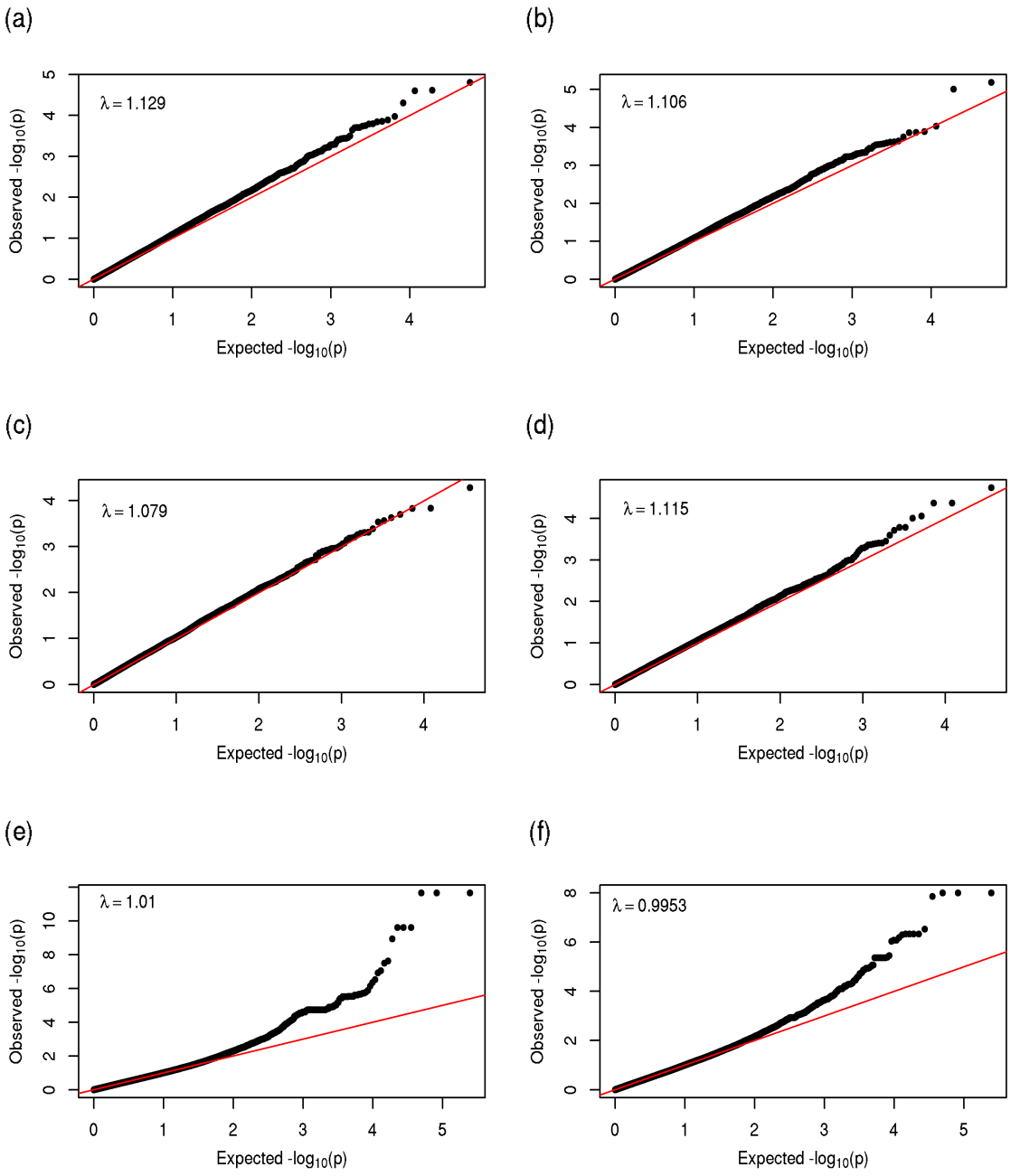

Figure S1. QQ plots for all single variant analyses by frequency class for SBP ((a): common, (c): low frequency, (e): rare) and DBP ((b): common, (d): low frequency, (f): rare). Here, singletons are included in the rare variant plots and calculations, although they were ultimately excluded from analyses.

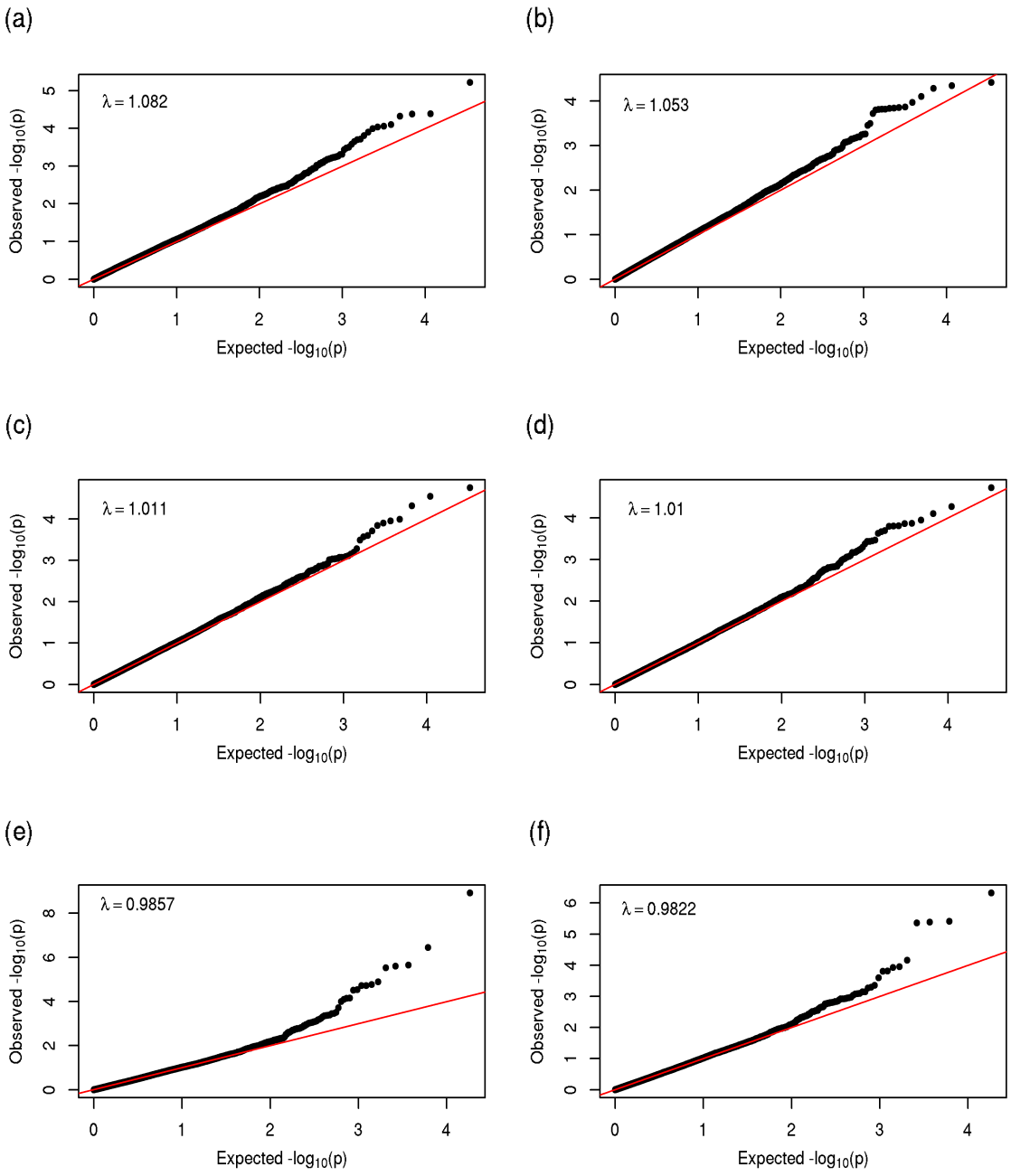

Figure S2. QQ plots for all gene-based analyses (SKAT, Burden-T1-all, and Burden-T1-del) for SBP ((a): SKAT, (c): Burden-T1-all, (e): Burden-T1-del), and DBP ((b): SKAT, (d): Burden-T1-all, (f): Burden-T1-del). All plots include genes containing just one variant; these were ultimately excluded from analyses.

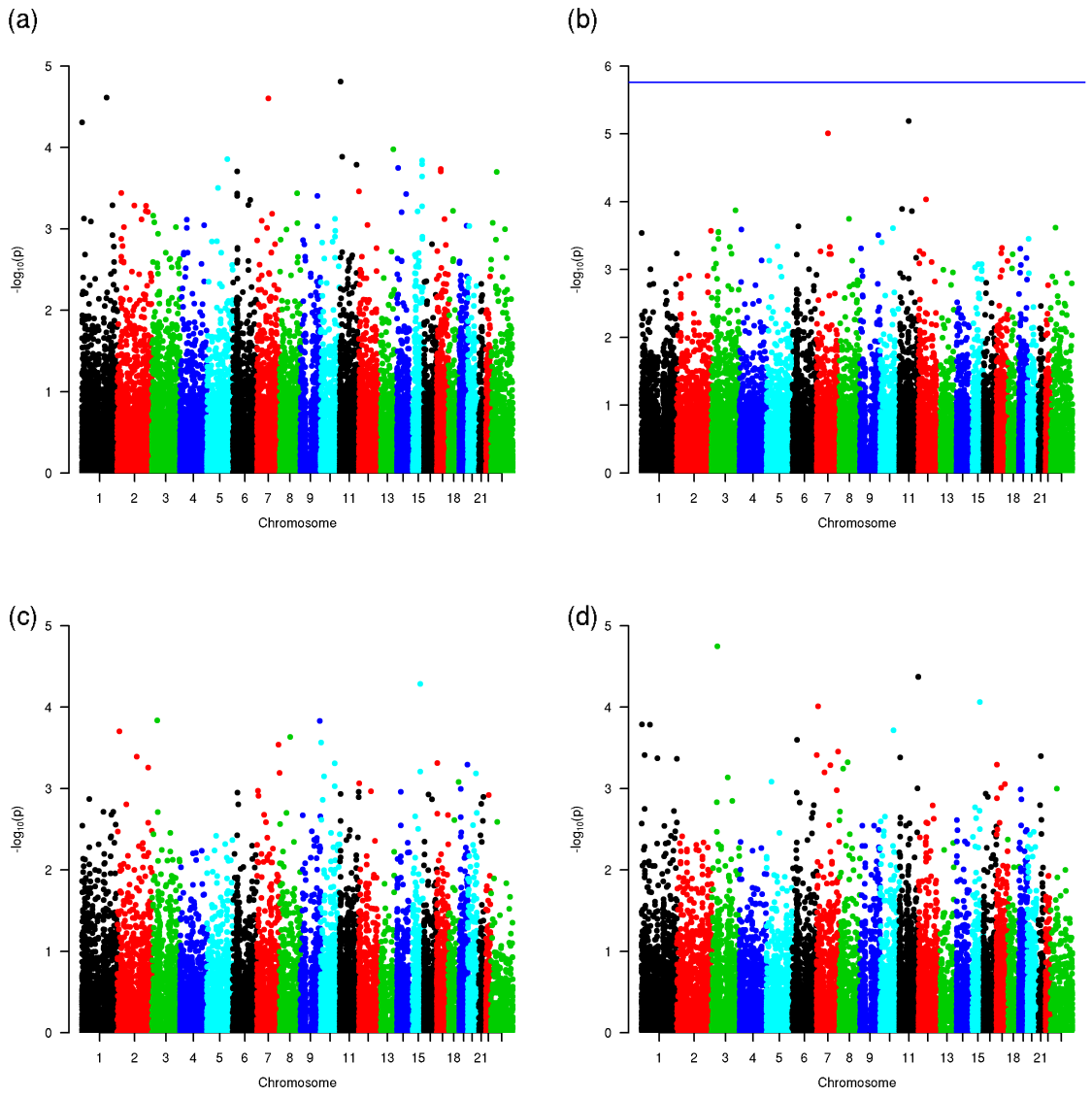

Figure S3. Manhattan plots for common (28,851 SNPs) and low frequency (17,985 SNPs) single variant analyses for SBP ((a): common, (c): low frequency) and DBP ((b): common, (d): low frequency). None of these variants reach statistical significance after Bonferroni correction.

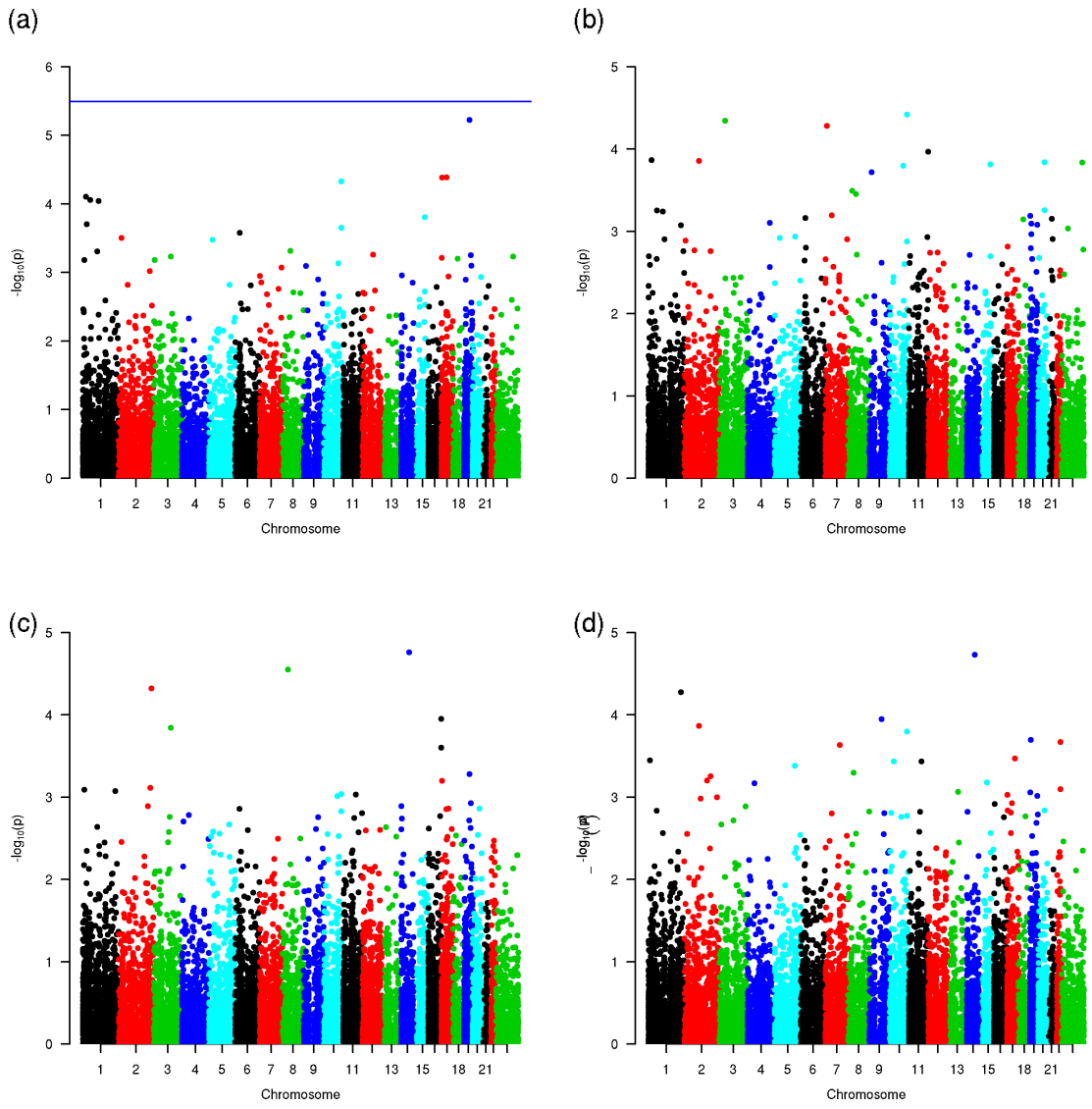

Figure S4. Manhattan plots for SKAT (15,554 genes) and Burden-T1-all (14,465 genes) analyses for SBP ((a): SKAT, (c): Burden-T1-all) and DBP ((b): SKAT, (d): Burden-T1-all). None of these genes reached statistical significance after Bonferroni correction.

Acknowledgements and Funding Sources (Complete)

We thank the participants of all of the studies involved that have made our analyses possible. This research was funded by grants from the NIH (R01 HL086694) under the aegis of the FEHGAS2 collaboration.

The LUC cohort study was supported by National Institutes of Health grant numbers R37-HL045508, R01-HL053353, R01-DK075787 and U01-HL054512. The authors acknowledge the assistance of the research staff and participants in Igbo-Ora and Ibadan, Oyo State, Nigeria.

The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Exome Chip genotyping was supported from grants R01-HL093029 and U01- HG004729 to MF. This manuscript has been reviewed and approved by CARDIA for scientific content.

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054464, HL054457, HL054481, HL087660, HL086694, HL119443) of the National Institutes of Health. Genotyping was performed at the Center for Inherited Disease Research (CIDR). We would also like to thank the families that participated in the GENOA study.

The HyperGEN cohort study is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah (Network Coordinating Center, Field Center, and Molecular Genetics Lab); University of Alabama at Birmingham (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin (Echo Genotyping Lab); Boston University (Field Center); University of Minnesota (Field Center and Biochemistry Lab); University of North Carolina (Field Center); Washington University (Data Coordinating Center); Weil Cornell Medical College (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html

The authors thank the participants in the AADM project, their families and their physicians. The crucial role of the study coordinators, as well as field and laboratory staff is gratefully acknowledged. The study was supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the National Institute of Minority Health and Health Disparities. The project was also supported by the NIDDK grant DK-54001.

HUFS was supported by NIH grants S06GM008016-320107 and S06GM008016-380111 from the NIGMS/MBRS/SCORE Program. Participant enrollment was carried out at the Howard University General Clinical Research Center, which is supported by grant 2M01RR010284 from the National Center for Research Resources, a component of the NIH. This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute in the Center for Research in Genomics and Global Health (CRGGH—Z01HG200362). CRGGH is also supported by National Institute of Diabetes and Digestive and Kidney Diseases, Center for Information Technology, and the Office of the Director at the NIH.

The Atherosclerosis Risk in Communities (ARIC) study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419).

The data analyses in this study, the use of data and samples from the GenNet project and the effort of the authors P.N., G.E.E. and A.C. were supported by NIH grant HL086694.

Chapter 3: Contributions of Rare Coding Variants in Hypotension Syndrome Genes to Population Blood Pressure Variation

Introduction

Hypertension (HTN) is a major risk factor for cardiovascular disease and affects ~30% of adults worldwide.[1,96] Recently, genome-wide association studies (GWAS) have identified common variants at ~166 loci associated with systolic blood pressure (SBP), diastolic blood pressure (DBP) and essential HTN, but these variants explain <5% of the phenotypic variance.[12–36] In contrast, rare variants in ~20 genes involved in renal salt handling and water balance have been implicated in monogenic forms of either hypertension or hypotension with electrolyte abnormalities.[10,53] However, the contribution of these genes to population level inter-individual variation in SBP and DBP is generally unknown. Since the existing GWAS loci do not include any of the 20 known hypertension or hypotension syndromic genes, it is reasonable to infer that common variation in these genes do not contribute greatly to inter-individual BP variation. The question is, why?

In 2008, Ji and colleagues[10] examined the effects of rare variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* on BP in the European ancestry Framingham Heart Study (FHS) subjects. The authors chose these ion channel genes because homozygotes for loss-of-function (LOF) variants in these diuretics targets lead to recessive renal salt-wasting hypotension syndromes (Bartter (prevalence ~1/1,000,000[10,97]) and Gitelman (prevalence ~1/40,000[10,98])), so that, consequently, LOF heterozygotes in these genes could reduce BP substantially and be protective against hypertension. Ji and colleagues identified 18

not previously validated and potentially LOF missense variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* in the FHS offspring cohort and demonstrated a significant protective effect (-6.3 mmHg for SBP and -3.4 mmHg for DBP) in their carriers in all ages between 25-60 years.

We attempted to also investigate *SLC12A3*, *SLC12A1*, and *KCNJ1* using identical methods but using whole exome sequence (WES) data on 7,810 European ancestry subjects from the population-based Atherosclerosis Risk in Communities (ARIC) cohort study. Like the FHS, ARIC is a longitudinal study with BP measurements over time and, therefore, allows for identical analyses as in the Ji et al. study. Success in such studies could motivate genome wide screens for genes with analogous effects.

Methods

Study Participants:

The ARIC (Atherosclerosis Risk in Communities) study is a population-based, prospective study on 15,792 individuals, including 11,478 European and 4,266 African ancestry US subjects from Forsyth County, NC, Jackson, MI, Minneapolis, MN and Washington County, MD.[60,61] There have been five visits, with individuals in the first visit (1987-1989) aged between 45 and 64 years. Subsequent visits occurred in 1990-1992, 1993-1995, 1996-1998, and 2011-2013. The data from the first four visits were used in this analysis. SBP and DBP were measured thrice at each of the first three visits, and twice at the fourth visit, using a random zero sphygmomanometer; the average of the (final) two measurements at each visit were used for analysis. In this study, exome sequence data from 7,444 European-Americans were analyzed (see below). All

participants provided written consent, and approval was obtained from the appropriate institutional review boards.

Whole Exome Sequencing, Variant Calling, and Quality Control:

ARIC samples were sequenced at the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC), as part of a larger set of CHARGE[99] samples. The exome sequencing protocol, variant calling, and quality control procedures are described elsewhere (Yu et al. *Circ Cardiovasc Genet*. 2015).[100] Among 14,443 CHARGE samples in the final cleaned set, which had mean depth of 78x coverage, there were 7,810 European American (EA) and 3,180 African American (AA) ARIC samples. Only the 7,810 EA samples were moved forward in our analysis at this point.

We then applied additional stringent filters to the data from these 7,810 individuals, and the following observations were culled: individuals with <90% call rate; variants with >10% missing genotype calls among samples; and variants failing Hardy Weinberg Equilibrium ($P < 1 \times 10^{-6}$). Finally, we used the genotype data to estimate the genetic relationships among the remaining 7,767 individuals using KING,[71] retaining only those individuals with the closest relationship being third degree. This provided a final dataset comprising 7,444 individuals.

Whole Genome Sequencing, Variant Calling, and Quality Control:

Genome sequencing was conducted, also as previously described in EA and AA ARIC study samples (Morrison et al. *Nat Genet*. 2013[101]; Morrison et al. *Am J Hum Genet*. 2017[102]). Briefly, individuals were sequenced on the HiSeq 2000 at 7.4x average

depth, and variants were called with goSNAP, which uses multiple variant callers including SNPTools, GATK, and GotCloud, in joint calling mode. Quality control procedures of the WGS data included the removal of variants failing Hardy-Weinberg Equilibrium ($P < 1 \times 10^{-14}$) per ancestry group, and those with site level inbreeding coefficient < -0.9, and principal components analysis to identify outliers in each ancestry group. The remaining 1,705 ARIC EA samples from these procedures[102] were analyzed in the present study.

Annotation to Identify Variants of Interest:

Annotation of variants in *SLC12A3*, *SLC12A1* and *KCNJ1*, with respect to their deleterious effects, was carried out using ANNOVAR with refGene annotations to identify the putative LOF missense variants ("nonsynonymous") in the selected transcripts used in the downstream analyses.

Phenotypes:

Longitudinal SBP and DBP phenotypes were calculated based on previously described methods,[10,103] using the first four visits from the ARIC study. Briefly, SBP and DBP were adjusted in a cubic regression on age within each age group (<35, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+ years) for the 7,444 individuals, and then used to adjust their measurements for those visits that were taken while on anti-hypertensive medications. While all SBP measurements were used, DBP measurements were restricted to those taken at age 55 years and below, as DBP is known to decline with increasing age beyond this point.[104] Subsequently, these residuals were adjusted for mean age in a linear

regression, separately by sex; in our study, this was done with all individuals with at least one visit, and regardless of the time span between the first and last visits. This differs from previously described methods in Levy et al. where the authors state that at least 4 examinations were required with a minimum time span of 10 years between the first and final examinations in the FHS study, and at least 3 examinations for FHS Offspring Cohort participants.[103] The resulting standardized residuals were used as phenotypes for our studies, and the mean of the non-standardized residuals among carriers of analyzed variants is presented as the effect size in Figures 1 and 2.

Transcripts, Orthologs and Paralogs Selection and Alignment:

Methods for ortholog and paralog selection were as previously described[10]; they are briefly restated here emphasizing pertinent differences. In the original study, the representative transcripts NM_000339 (*SLC12A3*), NM_000338 (*SLC12A1*), and NM_000220/NM_153764-7 (*KCNJ1*) were analyzed; in our study, we analyzed the same transcripts for *SLC12A3* and *SLC12A1*, and analyzed the transcript NM_153766 for *KCNJ1*. This *KCNJ1* transcript encodes an identical protein sequence to those encoded by other transcript variants NM_153764 and NM_153767, and presented with maximum positional similarity through multiple alignment to variants listed as conserved in Ji et al., as compared to the canonical transcript, NM_000220.

The orthologs analyzed in the Ji et al. study for *SLC12A3* were from human, mouse, rat, rabbit, dog, cow, chicken, zebrafish and winter flounder. In our study, the obsolete record XP_871112 from cow was replaced with NP_001193107. The *SLC12A1* orthologs analyzed in their study were from human, mouse, rat, rabbit, dog, chicken,

zebrafish and *Tetraodon*. In our study, NP_062007 from rat was removed from Refseq records as it is a nonsense-mediated mRNA decay (NMD) candidate, and XP_850426 from dog was removed from Refseq records as part of standard genome annotation processing (SGAP). These orthologs were dropped in our study. Finally, the orthologs studied for *KCNJ1* in their study were from human, mouse, rat, dog, chicken, cow, frog, zebrafish, fugu, *C. elegans*, sea urchin and *D. melanogaster*. In our study, XP_546403 from dog and XP_425795 from chicken were dropped due to their removal from Refseq records for SGAP, XP_684541 from zebrafish was replaced with the updated record NP_001092204, XP_585917 from cow was replaced with the updated record NP_0011179136, and NP_72245 from human was replaced with NP_722450 (corresponding to the selected transcript for *KCNJ1*).

The paralogs analyzed for *SLC12A3* and *SLC12A1* (*SLC12A* family) in both studies were from *SLC12A1*, *SLC12A3*, and additionally, to *SLC12A2*, *SLC12A4*, *SLC12A5*, *SLC12A6*, and *SLC12A7*. The paralogs analyzed for KCNJ1 (*KCNJ* family) in their study include *KCNJ2*, *KCNJ3*, *KCNJ4*, *KCNJ5*, *KCNJ6*, *KCNJ8*, *KCNJ9*, *KCNJ10*, *KCNJ11*, *KCNJ12*, *KCNJ13*, *KCNJ14*, *KCNJ15*, and *KCNJ16*. In our study, as for the orthologs, we used the human protein NP_72250 in accordance with the *KCNJ1* transcript analyzed, and additionally, NP_733838 from *KCNJ14* was removed due to inadequate support for its transcript, NM_170720.1 [http://www.ncbi.nlm.nih.gov/nuccore/].

Multiple alignments were performed with ClustalW v2.1[105,106] ([105,106]) with default parameters for all orthologs per gene, and all paralogs per gene family (one for the *KCNJ* family and one for the *SLC12A* family), with five alignments in total. This

differs slightly from the original alignment procedure described in Ji et al., in which it was stated that pairwise alignments were used in addition to multiple alignments to determine conserved positions. While the multiple pairwise alignments may have some differences to a multiple alignment, this difference is not expected to lead to any significant change in our downstream assessment of overall conservation of residues.

Selection Criteria and "Validation" Annotation:

The criteria used by Ji et al.[10] to identify putative LOF missense variants included: 1) allele frequency < 0.001, and 2) complete conservation across the selected orthologs. The exceptions to this were variants that were conserved in orthologs, but with the mutant residue present in paralogs, as these were thought to be sustainable within the species. Also, as in their study, we further annotated the analyzed variants from ARIC and the SNVs from FHS using PANTHER[107] (v9.0), SIFT[47] (with GRCh37/Ensembl 63) and PolyPhen-2[108] to determine the deleterious effects of our selected variants.

Statistical Analyses:

As all individuals in this study were unrelated, or related more remotely than third degree relatives. Standard one-tailed t-tests, under the assumption that the alternate alleles are BP-lowering, were carried out on standardized residuals as final phenotype values to compare carriers with non-carriers, assuming equal variance for both groups. Non-carriers were defined as those who have non-missing genotypes for all 65 variants; those with missing genotypes for any of the 65 were excluded from all calculations.

Protein Plots

Protein plots in Figures S1-S3 were created with the Protter software.[109]

Results

We examined rare missense variants in 7,444 EA ARIC subjects with WES data (no nonsense variants ultimately met the filtering criteria of frequency or conservation). The phenotypic (BP) and risk factor (age, sex, BMI) characteristics for these individuals are shown in Table 1. We identified a total of 216 missense variants (rate of 0.029 per individual) in the cleaned exome sequence data for *SLC12A3*, *SLC12A1* and *KCNJ1*. To assess their properties, we aligned these genes to their orthologs and paralogs as described in the Methods section with ClustalW2, and discovered conservation (protein sequence identity) of 41.5%, 40.6%, and 18.3% of residues across orthologs of *SLC12A3*, *SLC12A1* and *KCNJ1,* respectively, across all species considered in each of the multiple alignments. Of these, 65 variants (46, 17 and 2 for *SLC12A3*, *SLC12A1* and *KCNJ1*, respectively*)* had minor allele frequencies (MAF) <0.001 in the 7,444 ARIC EA samples. These occurred at highly sequenced bases (in 7,810 ARIC EA individuals: median depth 75X; range: 27-196X) demonstrating the high quality of the data.

We annotated these 65 variants with SIFT, PANTHER and Poly-Phen2, which confirmed that the vast majority (57, or 86.2%) of these variants were predicted to be pathogenic by all three programs (Table S1). We also used these versions of the software programs to obtain updated annotations for the 28 FHS SNVs, enabling comparisons between the two sets of variants (Table S2). All but three variants (89.3%) were predicted to be pathogenic by all three programs. Further, allele frequencies of these 65 variants in

non-Finnish Europeans from the Exome Aggregation Consortium (ExAC) are shown in Table S3, and those for the 28 FHS SNVs are presented in Table S4. There are 40 out of 65 ARIC variants (~61.5%) and 17 out of 28 FHS SNVs (60.7%) with nonzero non-Finnish ExAC allele frequencies, providing similar evidence for the existence of these genotype calls across the two studies. Further, examining the variant distributions across the three genes (Figures S1-S3) shows them to be evenly distributed across the domains in both studies.

Longitudinal SBP and DBP values were then analyzed by standard one-tailed t-tests to test for differences between carriers and non-carriers of the variants. The standardized residual phenotype values of carriers, modeled after Figure 2 in Ji et al., are shown in Figures 1 and 2 for SBP and DBP, respectively. The mean effect was -0.745 mmHg for SBP (t-test p=0.36) considering all 65 variants in 121 carriers and 6,750 non-carriers, and 1.177 mmHg for DBP (t-test p=0.83) considering 39 variants in 62 carriers and 3,478 non-carriers. As only DBP measurements under the age of 55 were analyzed in this study (see Methods), there were fewer variants in fewer individuals analyzed. Thus, these results neither show a prominent direction of effect, nor are they statistically significant.

We had whole genome sequence data (WGS) available for a subset of these individuals and variants, enabling comparisons of the same genotypes at *SLC12A3*, *SLC12A1* and *KCNJ1*. Genotype counts from the WES data were also present in the WGS data of 14 variants; seven variants were called in the same eight individuals in both sets concordantly. As this is too sparse to be truly informative, we compared the overall concordance of all genotype calls between the WES and WGS data for bi-allelic SNVs

from all 1,497 individuals. These comparisons had between 545,249 – 604,016 sites with a total concordance rate of 0.9995 (Table 2). As indicated in Table 2, there were many more variants, namely heterozygous variants, called in the exome sequence data than in the genome sequence data within this set.

We also examined a set of 10 variants from the 28 FHS SNVs that were also present in the ARIC exome sequencing dataset. Of these, 9 met the conservation and MAF criteria within ARIC, while 1 did not meet the conservation criterion (the proline at position 348 in *SLC12A1* has an alanine substitution in *Danio rerio*). Standardized residuals for carriers of all 10 variants are shown for SBP, and for carriers of the subset of 8 variants with DBP measurements available under age 55, in Figures 3 and 4, respectively. These results replicate those for the 30 variants in Ji et al. with similar mean effect sizes among carriers (SBP: -6.888 mmHg, p=0.02; DBP: -3.120 mmHg, p=0.11).

Discussion

We attempted to apply the criteria from Ji et al. to ARIC exome sequencing data to identify putative LOF variants within the Bartter and Gitelman syndromes *SLC12A3*, *SLC12A1*, and *KCNJ1* genes. While we failed to demonstrate the robustness of the methods used in Ji et al. in the ARIC exome sequencing as we did not see the BP-lowering effect they did when applying their filtering criteria to the ARIC exome sequencing data, we did replicate the LOF effect of several specific rare variants analyzed in their study, demonstrating evidence that such variants in these genes are protective against essential hypertension. These results merit further discussion. Rare LOF alleles at *SLC12A3*, *SLC12A1*, and *KCNJ1* do decrease blood pressure and hypertension risk in

heterozygotes in the general population but the specific alleles showing such effects are only a subset of all annotated deleterious alleles.

It is usually expected that identical computational genomic analyses should return similar, if not identical, results; however, many inadvertent differences can occur. First, there are many differences in the sequence versions of the orthologs and paralogs used in the two studies because many records were either updated or dropped from the NCBI Refseq records over the past 9 years. We had 18-42% sites conserved across all species in the multiple alignments, while the authors in Ji et al. state 18-24% conservation rate for each of the three genes. Sequence and ClustwalW software version differences, as well as differences in calculation methods of these values, are likely to have led to some discrepancies, even for *SLC12A3*, for which we used the same orthologs and paralogs. In our multiple alignments, 21 out of 28 FHS SNVs are completely conserved, of which 17 are listed in Table 1 of Ji et al. as completely conserved in their study (the remaining four are listed as conserved in vertebrates in their study). Additionally, dropping sequence records that were obsolete by the time we conducted this study likely contributed to the higher percentages of conserved sites in the species studied, though these positions still represent a conserved set.

The two studies also used different variant detection methods: we used exome sequencing whereas Ji et al. used temperature gradient capillary electrophoresis (TGCE) with confirmation by PCR amplification and Sanger sequencing from the original DNA sample. The original study stated a "high" sensitivity of detection, having identified known single nucleotide variants (SNVs) in FHS, buttressed by a high concordance with frequencies in previous studies. Follow-up studies state that the vast majority (14 out of

18, with conflicting results for an addition two variants) of the variants in their study were shown in *Xenopus laevis* oocytes and HEK293 cells to be LOF.[110–112] We use WES data in this study, at high sequencing depths that are in the range of previous estimates deemed sufficient for detection of the vast majority of heterozygous variants.[113–115] The WGS data used for concordance comparison indicated some discordance with the WES data, but are low-coverage and therefore, it is likely that variants were missed in these data that were called in the WES data. However, it should also be noted that WES and WGS data often contain both false positives and false negatives,[116,117] despite a low error rate, for the rarest of variants. Our annotation of the 65 analyzed ARIC variants and the 28 FHS SNVs analyzed in Ji et al. with the software programs SIFT, PANTHER, and PolyPhen support the predicted pathogenicity of both variant sets, and similar fractions of variants in both sets were present in the ExAC non-Finnish European population, demonstrating high concordance with their properties. However, it is not necessarily surprising for sequencing errors to also display this profile of pathogenicity due to the increased probability that rare variants are sequencing errors and that the analyzed variants were specifically chosen at sites preselected for their conservation among a set of species.

Despite these differences, the carriers of the subset of variants from the Ji et al. study that were also in the ARIC WES data present similar effect sizes for SBP (10 variants) and DBP (eight variants) as in the Ji et al. study, and nearly all passed the same selection criteria in ARIC. As the distributions of variants in both studies seem to show no distinct or study-specific patterns or differences, it remains possible that only certain LOF variants at mechanistically specific locations are essential for BP regulation. This

replication is especially notable because the FHS offspring cohort is younger in their visits as compared to ARIC individuals, who were minimum 45 years of age at baseline, and the age difference affects the proportion of individuals whose measurements necessitated adjustment for medication use. Further, the methods detailed in Levy et al. indicate that the effects reflect a longer time span of measurements, as they required samples to have at least 3-4 visits over at least 10 years, while our samples had a maximum of four visits over 12 years. These differences can affect analyses, but in this case the effect is still visible. Additionally, this represents a replication of the effects of rare variants with very low minor allele counts, which demonstrates the utility of exome sequencing data for study of such rare variants, though greater study is required to determine more appropriate filtering methods.

In summary, our attempt to replicate the methods in the Ji et al. study to detect rare and potentially LOF variants in *SLC12A3*, *SLC12A1*, and *KCNJ1* reducing SBP and DBP in variant carriers as compared to non-carriers using WES data was overall unsuccessful. Although the two studies are comparable in numerous ways there are also pertinent differences that can lead to the discrepant outcomes. Regardless of this, we successfully replicated the reduction effects in SBP and DBP with a subset of variants from the Ji et al. study that were present in the ARIC study, which upholds the use of such sequencing data for the study of very rare variants and that *SLC12A3*, *SLC12A1*, and *KCNJ1* are indeed genes protective in HTN. Additionally, this type of analysis demonstrates the utility of this method for analyzing specific rare variants in aggregate. Though hypertension is a common disease, as the particular renal salt wasting syndromes of interest in this study (Bartter and Gitelman) are rare, the method is applicable for

identifying rare LOF variants LOF in other rare diseases. Nevertheless, our study also supports the idea that software prediction and annotation of rare, putatively deleterious variants may be insufficient for detecting disease genes, and that there must be appropriate variant filtering methods, and subsequent functional analyses and validation of these variants.

Acknowledgements

The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their important contributions. Funding support for "Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). DNA sequencing for this study was carried out at the Baylor College of Medicine Human Genome Sequencing Center and supported by the National Human Genome Research Institute grants U54 HG003273 and UM1 HG008898. This study was funded by NIH grant HL086694 to A. C.

Chapter 3 Figures

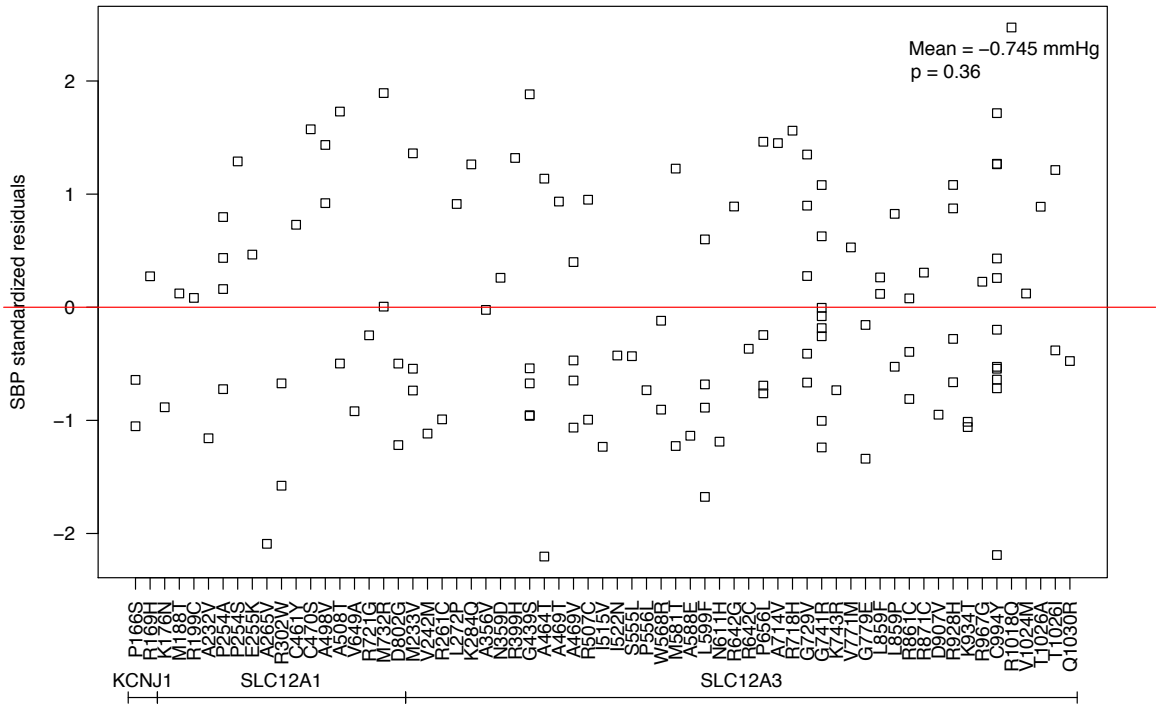

Figure 1. Age- and sex-adjusted SBP residuals for 121 carriers of 65 variants, depicting the mean effect of the carriers.

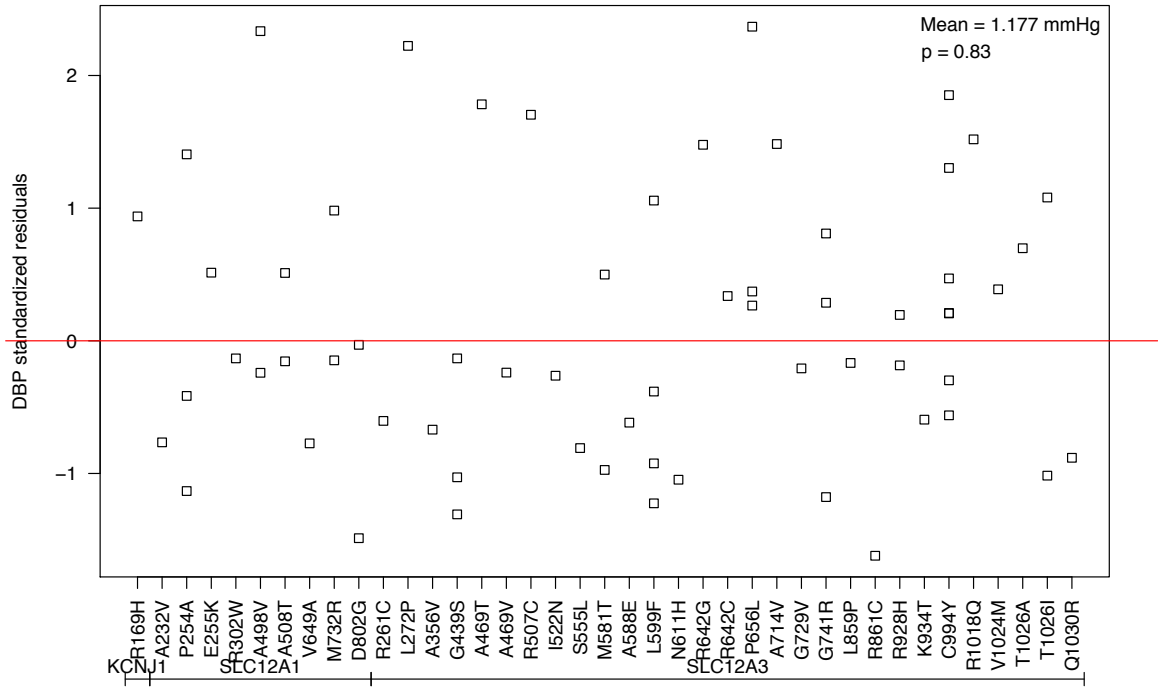

Figure 2. Age- and sex-adjusted DBP residuals for 62 carriers of 39 variants, depicting the mean effect of the carriers.

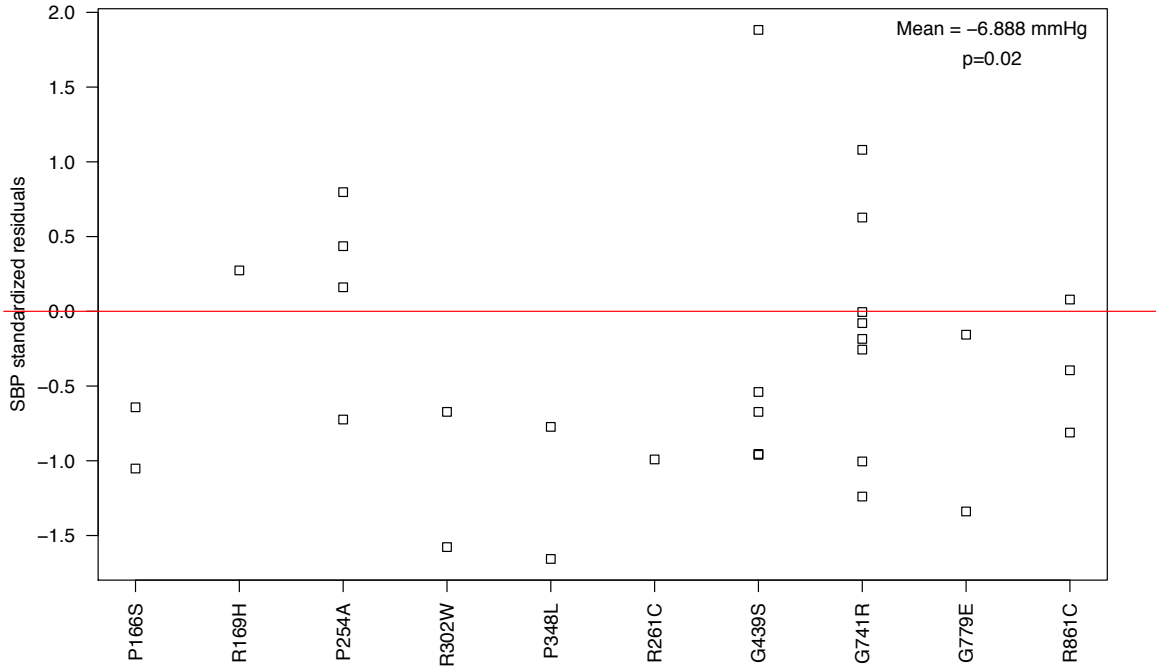

Figure 3. Age- and sex-adjusted SBP residuals for 10 FHS variants in ARIC, depicting the mean effect of the carriers.

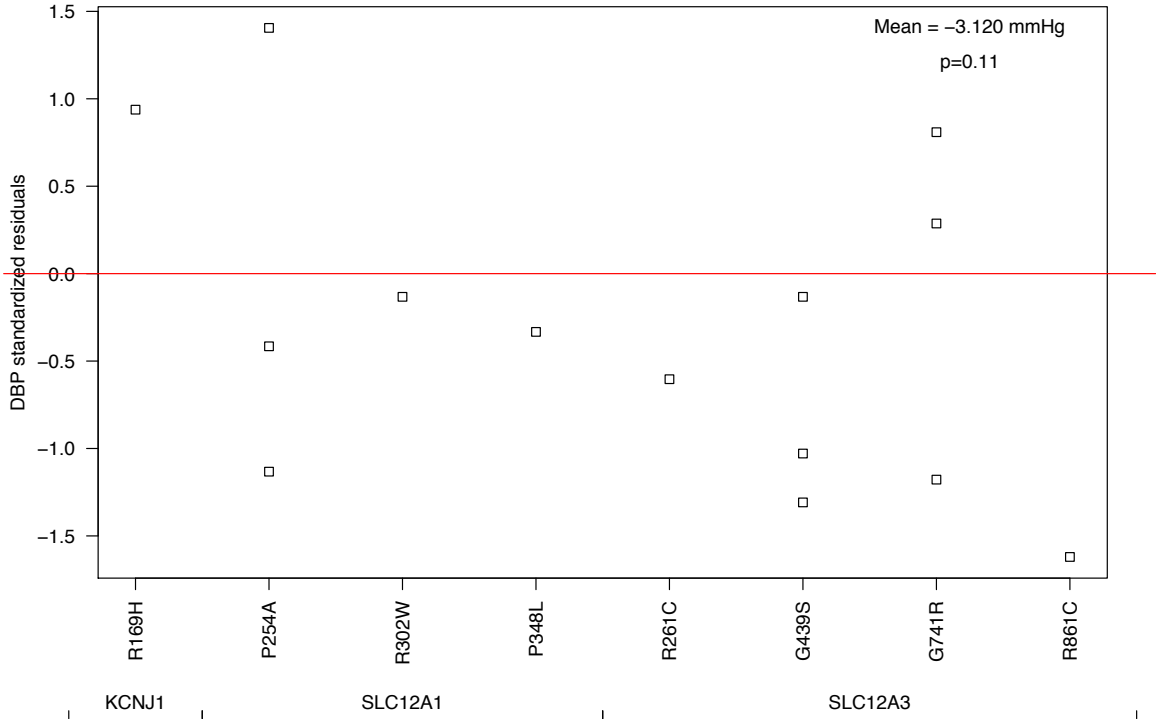

Figure 4. Age- and sex-adjusted DBP residuals for 8 FHS variants in ARIC, depicting the mean effect of the carriers.

Chapter 3 Tables

Table 1. Sample characteristics of 7,444 ARIC subjects analyzed in this study.

Visit/ Phase	Age (SD)	BMI (SD)	SBP unadjusted (SD)	SBP adjusted (SD)	DBP unadjusted (SD)	DBP adjusted (SD)	% medicated
1	54.288 (5.665)	26.967 (4.815)	118.204 (16.487)	121.848 (19.649)	71.765 (9.678)	73.055 (10.416)	18.8
2	57.223 (5.64)	27.285 (4.887)	119.754 (17.299)	124.107 (21.135)	71.356 (9.698)	73.096 (10.379)	20.6
3	60.289 (5.609)	27.91 (5.141)	122.597 (17.696)	128.491 (22.436)	70.779 (9.932)	74.032 (10.908)	25.6
4	63.095 (5.592)	28.25 (5.209)	125.916 (18.187)	133.533 (23.517)	70.048 (9.823)	73.092 (10.587)	30.9

Of 7,444 individuals (53.76% female), SBP residuals were calculated for 7,443 and DBP residuals were calculated for 3,835 subjects.

Table 2. Whole exome (WES) versus whole genome (WGS) sequencing genotype call comparisons for all calls in 1,497 individuals.*

		# WES genotype calls		
		0	**1**	**2**
# WGS calls	**0**	870,939,675	238,845	343
	1	134,332	13,030,169	33,062
	2	104	44371	9,052,661

0, homozygous reference genotype; 1, heterozygous genotype; 2, homozygous alternate genotype

*Comparisons of non-missing genotype calls across individuals in both datasets; the overall genotype concordance is 0.9995. The calculation is inclusive of homozygous reference calls for variants called in other samples not within this subset of 1,497 as the samples included in this study were called as part of a larger set of individuals (see Methods).

Chapter 3 Supplement

Supplementary methods, tables, and figures

Supplementary Methods

Sequences of orthologs and paralogs used in this study:

NCBI links for orthologs/paralogs may be found at http://www.ncbi.nlm.nih.gov/protein/.

GenBank Protein IDs for the human are: *KCNJ1* (NP_722450.1), *SLC12A1* (NP_000329.2), and *SLC12A3* (NP_000330.2).

Orthologs for *KCNJ1* include: *Rattus norvegicus* (NP_058719.1), *Mus musculus* (NP_062633.1), *Drosophila melanogaster* (NP_651131.2), *Xenopus laevis* (AAH79788.1), *Strongylocentrotus purpuratus* (XP_789112.1), *Takifugu rubripes* (ABB87033.1), *Danio rerio* (NP_001092204.1), *Bos taurus* (NP_001179136.1), *Caenorhabditis elegans* (NP_509138.2).

Paralogs for human KCNJ family include: *KCNJ2* (NP_000882.1), *KCNJ3* (NP_002230.1), *KCNJ4* (NP_004972.1), *KCNJ5* (NP_000881.3), *KCNJ6* (NP_002231.1), *KCNJ8* (NP_004973.1), *KCNJ9* (NP_004974.2), *KCNJ10* (NP_002232.2), *KCNJ11* (NP_000516.3), *KCNJ12* (NP_066292.2), *KCNJ13* (NP_002233.2), *KCNJ15* (NP_733933.1), *KCNJ16* (NP_733938.2)

Orthologs for *SLC12A1* include: *Tetraodon nigroviridis* (CAF99849.1), *Danio rerio* (NP_001002080.1), *Mus musculus* (NP_899197.2), *Gallus gallus* (XP_413814.3)

Orthologs for *SLC12A3* include: *Oryctolagus cuniculus* (AAC33139.1), *Pseudopleuronectes americanus* (AAL26926.1), *Danio rerio* (NP_001038545.1), *Rattus norvegicus* (NP_062218.3), *Mus musculus* (NP_062288.2), *Bos taurus* (NP_001193107.1), *Canis lupus familiaris* (XP_535292.3), *Gallus gallus* (XP_414059.4)

Paralogs for human SLC12A family include: *SLC12A2* (NP_001037.1), *SLC12A4* (NP_005063.1), *SLC12A5* (NP_065759.1), *SLC12A6* (NP_005126.1), *SLC12A7* (NP_006589.2).

Table S1. ARIC variant annotation of deleterious effects using SIFT, PolyPhen-2 and PANTHER cSNP software.

Gene	Substitution	Variant	SIFT.pred	PPhen.pred	PANTHER.pred
KCNJ1*	R169H	11:128709633:G:A	DAMAGING	probably damaging	probably damaging
KCNJ1*	P166S	11:128709643:C:T	DAMAGING	benign	probably damaging
SLC12A1	K176N	15:48512938:G:C	DAMAGING	probably damaging	probably damaging
SLC12A1	M188T	15:48513128:T:C	DAMAGING	probably damaging	probably damaging
SLC12A1	R199C	15:48513160:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	A232V	15:48518739:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	P254A	15:48521421:C:G	DAMAGING	probably damaging	probably damaging
SLC12A1	P254S	15:48521421:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	E255K	15:48521424:G:A	DAMAGING	probably damaging	probably damaging
SLC12A1	A265V	15:48521455:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	R302W	15:48522629:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	C461Y	15:48537031:G:A	DAMAGING	probably damaging	probably damaging
SLC12A1	C470S	15:48537057:T:A	DAMAGING	probably damaging	probably damaging
SLC12A1	A498V	15:48539146:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	A508T	15:48539175:G:A	DAMAGING	probably damaging	probably damaging
SLC12A1	V649A	15:48548011:T:C	DAMAGING	probably damaging	probably damaging
SLC12A1	R721G	15:48559764:C:G	TOLERATED	benign	probably damaging
SLC12A1	M732R	15:48559798:T:G	TOLERATED	benign	probably damaging
SLC12A1	D802G	15:48566770:A:G	DAMAGING	benign	probably damaging
SLC12A3	M233V	16:56904103:A:G	DAMAGING	possibly damaging	probably damaging
SLC12A3	V242M	16:56904130:G:A	TOLERATED	probably damaging	probably damaging
SLC12A3	R261C	16:56904577:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	L272P	16:56904611:T:C	DAMAGING	probably damaging	probably damaging
SLC12A3	K284Q	16:56904646:A:C	DAMAGING	probably damaging	probably damaging
SLC12A3	A356V	16:56906670:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	N359D	16:56906678:A:G	DAMAGING	probably damaging	probably damaging
SLC12A3	R399H	16:56913000:G:A	DAMAGING	probably damaging	probably damaging

SLC12A3	G439S	16:56913119:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	A464T	16:56913508:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	A469T	16:56913523:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	A469V	16:56913524:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	R507C	16:56914117:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	I515V	16:56914141:A:G	DAMAGING	possibly damaging	probably damaging
SLC12A3	I522N	16:56914163:T:A	DAMAGING	probably damaging	probably damaging
SLC12A3	S555L	16:56916404:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	P556L	16:56916407:C:T	TOLERATED	possibly damaging	probably damaging
SLC12A3	W568R	16:56917993:T:C	DAMAGING	probably damaging	probably damaging
SLC12A3	M581T	16:56918033:T:C	DAMAGING	probably damaging	probably damaging
SLC12A3	A588E	16:56918054:C:A	DAMAGING	probably damaging	possibly damaging
SLC12A3	L599F	16:56918086:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	N611H	16:56919182:A:C	DAMAGING	probably damaging	probably damaging
SLC12A3	R642G	16:56919275:C:G	DAMAGING	probably damaging	probably damaging
SLC12A3	R642C	16:56919275:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	P656L	16:56920317:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	A714V	16:56920968:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	R718H	16:56920980:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	G729V	16:56921844:G:T	DAMAGING	probably damaging	probably damaging
SLC12A3	G741R	16:56921879:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	K743R	16:56921886:A:G	DAMAGING	probably damaging	probably damaging
SLC12A3	V771M	16:56924211:G:A	DAMAGING	benign	possibly damaging
SLC12A3	G779E	16:56924236:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	L859F	16:56928469:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	L859P	16:56928470:T:C	DAMAGING	probably damaging	probably damaging
SLC12A3	R861C	16:56928475:C:T	TOLERATED	probably damaging	probably damaging
SLC12A3	R871C	16:56928505:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	D907V	16:56933501:A:T	DAMAGING	probably damaging	probably damaging
SLC12A3	R928H	16:56936320:G:A	DAMAGING	probably damaging	probably damaging

SLC12A3	K934T	16:56936338:A:C	TOLERATED	benign	probably damaging
SLC12A3	R967G	16:56938322:A:G	DAMAGING	probably damaging	probably damaging
SLC12A3	C994Y	16:56947205:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	R1018Q	16:56947277:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	V1024M	16:56947294:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	T1026A	16:56947300:A:G	DAMAGING	probably damaging	probably damaging
SLC12A3	T1026I	16:56947301:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	Q1030R	16:56947313:A:G	DAMAGING *Warning! Low confidence.	probably damaging	probably damaging

*Although the protein sequence for NP_722450.1 (*KCNJ1*) was used as input to PolyPhen-2, the software annotated *KCNJ1* variants in this transcript with equivalent positions in the canonical transcript NM_000220 (NP_000211.1).

Table S2. FHS variant annotation of deleterious effects using SIFT, PolyPhen-2 and PANTHER cSNP software.

Gene	Substitution	Variant	SIFT.pred	PPhen.pred	PANTHER.pred
KCNJ1*	H251Y	11:128709388:C:T	DAMAGING	probably damaging	probably damaging
KCNJ1*	R193P	11:128709561:G:C	DAMAGING	probably damaging	probably benign
KCNJ1*	R169H	11:128709633:G:A	DAMAGING	probably damaging	probably damaging
KCNJ1*	P166S	11:128709643:C:T	DAMAGING	benign	probably damaging
SLC12A1	T235I**	15:48518748:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	P254A	15:48521421:C:G	DAMAGING	probably damaging	probably damaging
SLC12A1	R302W	15:48522629:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	P348L	15:48524991:C:T	DAMAGING	probably damaging	probably damaging
SLC12A1	N399S	15:48527182:A:G	DAMAGING	probably damaging	probably damaging
SLC12A1	L505V	15:48539166:C:G	DAMAGING	probably damaging	probably damaging
SLC12A1	P569H	15:48541793:C:A	DAMAGING	probably damaging	probably damaging
SLC12A1	Y1070C	15:48594991:A:G	DAMAGING	probably damaging	probably damaging
SLC12A1	P1083A	15:48595029:C:G	DAMAGING	probably damaging	probably damaging
SLC12A3	E112X	16:56901033:G:T	N/A	NA	possibly damaging
SLC12A3	L155F	16:56902242:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	S188F	16:56903698:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	A232T	16:56904100:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	R261C	16:56904577:C:T	DAMAGING	probably damaging	probably damaging
SLC12A3	R399L	16:56913000:G:T	DAMAGING	probably damaging	probably damaging
SLC12A3	G439S	16:56913119:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	F495L	16:56914083:C:A	DAMAGING	possibly damaging	probably damaging
SLC12A3	P560H	16:56917970:C:A	DAMAGING	probably damaging	probably damaging
SLC12A3	G613S	16:56919188:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	G741R	16:56921879:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	G779E	16:56924236:G:A	DAMAGING	probably damaging	probably damaging
SLC12A3	R861C	16:56928475:C:T	TOLERATED	probably damaging	probably damaging

SLC12A3	R964Q	16:56938314:G:A	DAMAGING	possibly damaging	probably damaging
SLC12A3	G989R	16:56947189:G:A	DAMAGING	probably damaging	possibly damaging

*Although the protein sequence for NP_722450.1 (*KCNJ1*) was used as input to PolyPhen-2, the software annotated *KCNJ1* variants in this transcript with equivalent positions in the canonical transcript NM_000220 (NP_000211.1).

**The variant annotated as T235M in FHS in the *SLC12A1* transcript was annotated here as T235I (c.704C>T)

Table S3. ARIC variant allele frequencies in non-Finnish European individuals from the Exome Aggregation Consortium (ExAC).

Gene	Substitution	Variant	ExAC_AF	ARIC_AF
KCNJ1	R169H	11:128709633:G:A	1.54×10^{-5}	6.72×10^{-5}
KCNJ1	P166S	11:128709643:C:T	1.08×10^{-4}	1.34×10^{-4}
SLC12A1	K176N	15:48512938:G:C	1.51×10^{-5}	6.72×10^{-5}
SLC12A1	M188T	15:48513128:T:C	7.49×10^{-5}	6.72×10^{-5}
SLC12A1	R199C	15:48513160:C:T	1.50×10^{-5}	6.72×10^{-5}
SLC12A1	A232V	15:48518739:C:T	4.51×10^{-5}	6.72×10^{-5}
SLC12A1	P254A	15:48521421:C:G	NA	2.69×10^{-4}
SLC12A1	P254S	15:48521421:C:T	NA	6.72×10^{-5}
SLC12A1	E255K	15:48521424:G:A	NA	6.72×10^{-5}
SLC12A1	A265V	15:48521455:C:T	3.39×10^{-5}	6.72×10^{-5}
SLC12A1	R302W	15:48522629:C:T	NA	1.34×10^{-4}
SLC12A1	C461Y	15:48537031:G:A	NA	6.72×10^{-5}
SLC12A1	C470S	15:48537057:T:A	NA	6.72×10^{-5}
SLC12A1	A498V	15:48539146:C:T	NA	1.34×10^{-4}
SLC12A1	A508T	15:48539175:G:A	9.02×10^{-5}	1.34×10^{-4}
SLC12A1	V649A	15:48548011:T:C	NA	6.72×10^{-5}
SLC12A1	R721G	15:48559764:C:G	NA	6.74×10^{-5}
SLC12A1	M732R	15:48559798:T:G	NA	1.35×10^{-4}
SLC12A1	D802G	15:48566770:A:G	1.80×10^{-4}	1.36×10^{-4}
SLC12A3	M233V	16:56904103:A:G	NA	2.02×10^{-4}
SLC12A3	V242M	16:56904130:G:A	1.52×10^{-5}	6.78×10^{-5}
SLC12A3	R261C	16:56904577:C:T	9.00×10^{-5}	6.72×10^{-5}
SLC12A3	L272P	16:56904611:T:C	NA	6.72×10^{-5}
SLC12A3	K284Q	16:56904646:A:C	4.65×10^{-5}	6.73×10^{-5}
SLC12A3	A356V	16:56906670:C:T	NA	6.74×10^{-5}
SLC12A3	N359D	16:56906678:A:G	NA	6.75×10^{-5}
SLC12A3	R399H	16:56913000:G:A	2.04×10^{-5}	6.72×10^{-5}
SLC12A3	G439S	16:56913119:G:A	4.11×10^{-4}	3.36×10^{-4}
SLC12A3	A464T	16:56913508:G:A	3.02×10^{-5}	1.35×10^{-4}
SLC12A3	A469T	16:56913523:G:A	2.87×10^{-4}	6.73×10^{-5}
SLC12A3	A469V	16:56913524:C:T	1.21×10^{-4}	2.69×10^{-4}
SLC12A3	R507C	16:56914117:C:T	2.48×10^{-4}	1.35×10^{-4}
SLC12A3	I515V	16:56914141:A:G	9.34×10^{-5}	6.77×10^{-5}
SLC12A3	I522N	16:56914163:T:A	NA	6.83×10^{-5}
SLC12A3	S555L	16:56916404:C:T	7.49×10^{-5}	6.72 x 10+
SLC12A3	P556L	16:56916407:C:T	NA	6.72×10^{-5}
SLC12A3	W568R	16:56917993:T:C	1.50×10^{-5}	1.34×10^{-4}
SLC12A3	M581T	16:56918033:T:C	2.10×10^{-4}	1.34×10^{-4}
SLC12A3	A588E	16:56918054:C:A	1.50×10^{-5}	6.72×10^{-5}
SLC12A3	L599F	16:56918086:C:T	1.80×10^{-4}	2.69×10^{-4}

SLC12A3	N611H	16:56919182:A:C	NA	6.76×10^{-5}
SLC12A3	R642G	16:56919275:C:G	NA	6.96×10^{-5}
SLC12A3	R642C	16:56919275:C:T	NA	6.97×10^{-5}
SLC12A3	P656L	16:56920317:C:T	2.26×10^{-4}	2.69×10^{-4}
SLC12A3	A714V	16:56920968:C:T	1.96×10^{-4}	6.72×10^{-5}
SLC12A3	R718H	16:56920980:G:A	1.51×10^{-5}	6.72×10^{-5}
SLC12A3	G729V	16:56921844:G:T	2.73×10^{-4}	3.37×10^{-4}
SLC12A3	G741R	16:56921879:G:A	6.34×10^{-4}	5.38×10^{-4}
SLC12A3	K743R	16:56921886:A:G	NA	6.73×10^{-5}
SLC12A3	V771M	16:56924211:G:A	NA	6.72×10^{-5}
SLC12A3	G779E	16:56924236:G:A	NA	1.34×10^{-4}
SLC12A3	L859F	16:56928469:C:T	NA	1.34×10^{-4}
SLC12A3	L859P	16:56928470:T:C	1.56×10^{-4}	1.34×10^{-4}
SLC12A3	R861C	16:56928475:C:T	1.85×10^{-4}	2.02×10^{-4}
SLC12A3	R871C	16:56928505:C:T	3.02×10^{-5}	6.72×10^{-5}
SLC12A3	D907V	16:56933501:A:T	NA	6.72×10^{-5}
SLC12A3	R928H	16:56936320:G:A	7.49×10^{-5}	2.69×10^{-4}
SLC12A3	K934T	16:56936338:A:C	1.50×10^{-5}	1.34×10^{-4}
SLC12A3	R967G	16:56938322:A:G	7.49×10^{-5}	6.74×10^{-5}
SLC12A3	C994Y	16:56947205:G:A	3.15×10^{-4}	7.39×10^{-4}
SLC12A3	R1018Q	16:56947277:G:A	1.50×10^{-5}	6.72×10^{-5}
SLC12A3	V1024M	16:56947294:G:A	3.00×10^{-5}	6.72×10^{-5}
SLC12A3	T1026A	16:56947300:A:G	NA	6.72×10^{-5}
SLC12A3	T1026I	16:56947301:C:T	1.50×10^{-5}	1.34×10^{-4}
SLC12A3	Q1030R	16:56947313:A:G	1.50×10^{-5}	6.72×10^{-5}

ExAC_AF, ExAC non-Finnish European allele frequency; ARIC_AF, ARIC EA allele frequency

Table S4. FHS SNV allele frequencies in non-Finnish European individuals from the Exome Aggregation Consortium (ExAC).

Gene	Substitution	Variant	ExAC_AF
KCNJ1	H251Y	11:128709388:C:T	1.50×10^{-5}
KCNJ1	R193P	11:128709561:G:C	NA
KCNJ1	R169H	11:128709633:G:A	1.54×10^{-5}
KCNJ1	P166S	11:128709643:C:T	1.08×10^{-4}
SLC12A1	T235M	15:48518748:C:T	NA
SLC12A1	P254A	15:48521421:C:G	NA
SLC12A1	R302W	15:48522629:C:T	NA
SLC12A1	P348L	15:48524991:C:T	1.50×10^{-5}
SLC12A1	N399S	15:48527182:A:G	NA
SLC12A1	L505V	15:48539166:C:G	1.50×10^{-5}
SLC12A1	P569H	15:48541793:C:A	1.50×10^{-5}
SLC12A1	Y1070C	15:48594991:A:G	NA
SLC12A1	P1083A	15:48595029:C:G	1.50×10^{-5}
SLC12A3	E112X	16:56901033:G:T	NA
SLC12A3	L155F	16:56902242:C:T	1.59×10^{-5}
SLC12A3	S188F	16:56903698:C:T	1.50×10^{-5}
SLC12A3	A232T	16:56904100:G:A	1.50×10^{-4}
SLC12A3	R261C	16:56904577:C:T	9.00×10^{-5}
SLC12A3	R399L	16:56913000:G:T	4.07×10^{-5}
SLC12A3	G439S	16:56913119:G:A	4.11×10^{-4}
SLC12A3	F495L	16:56914083:C:A	NA
SLC12A3	P560H	16:56917970:C:A	NA
SLC12A3	G613S	16:56919188:G:A	NA
SLC12A3	G741R	16:56921879:G:A	6.34×10^{-4}
SLC12A3	G779E	16:56924236:G:A	NA
SLC12A3	R861C	16:56928475:C:T	1.85×10^{-4}
SLC12A3	R964Q	16:56938314:G:A	2.70×10^{-4}
SLC12A3	G989R	16:56947189:G:A	7.49×10^{-5}

ExAC_AF, ExAC non-Finnish European allele frequency

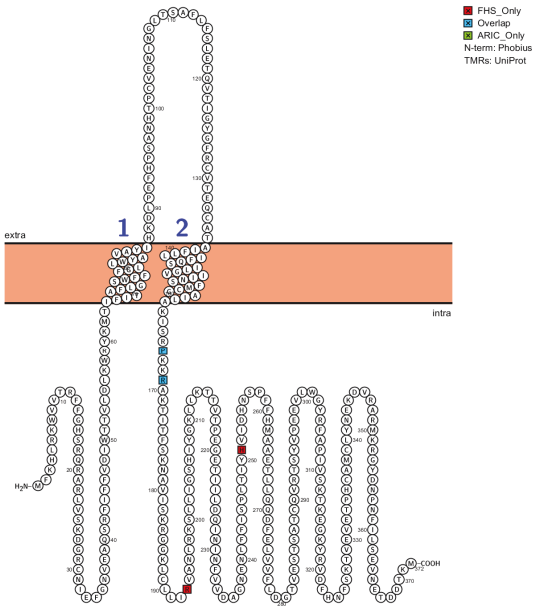

Figure S1. Distribution of *KCNJ1* variants in FHS and ARIC. SNVs unique to FHS are shown in red, SNVs unique to ARIC in green, and overlapping variants in blue.

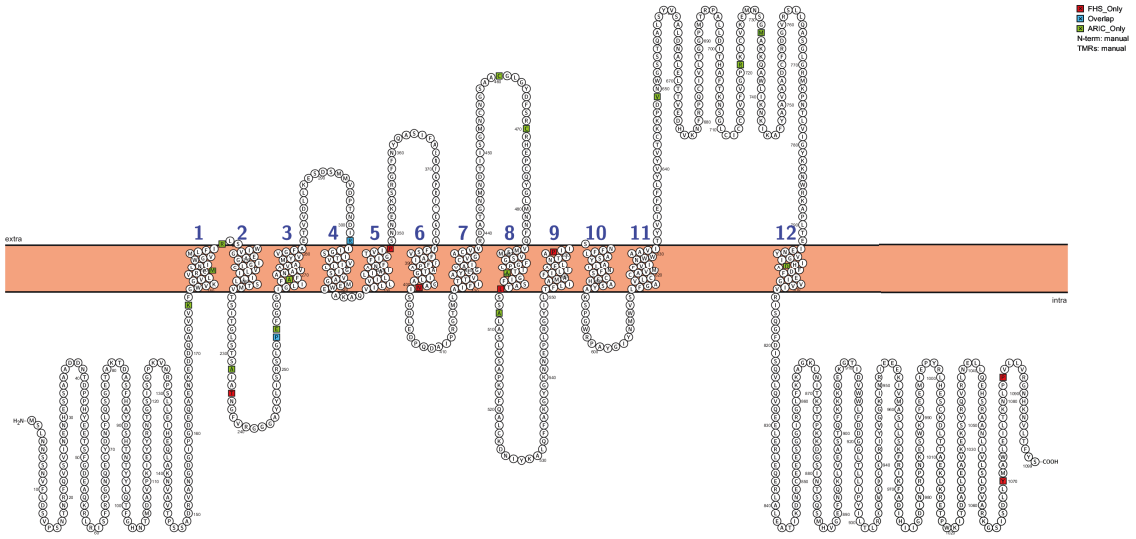

Figure S2. Distribution of *SLC12A1* variants in FHS and ARIC. SNVs unique to FHS are shown in red, SNVs unique to ARIC in green, and overlapping variants in blue.

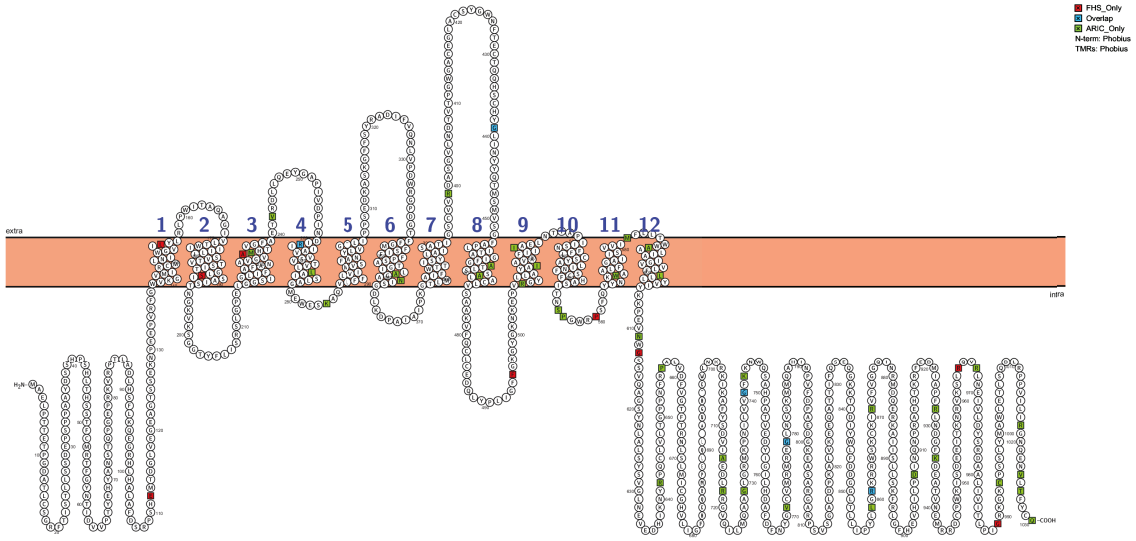

Figure S3. Distribution of *SLC12A3* variants in FHS and ARIC. SNVs unique to FHS are shown in red, SNVs unique to ARIC in green, and overlapping variants in blue.

Chapter 4: Analysis of putative cis-regulatory elements regulating blood pressure variation

Introduction

Several genome-wide association studies (GWAS) have revealed associations of common variants at hundreds of loci with inter-individual variation in blood pressure (BP)[12–36]; however, the specific causal genes and variants at most of these loci have yet to be determined. Consequently, the relative roles of coding and non-coding causal variation affecting BP remain unknown. In addition to these loci, a set of 20 genes involved in kidney salt transport and related mechanisms have been studied in the context of renal physiology and BP, because loss-of-function protein coding defects within them have been previously demonstrated to result in rare monogenic BP syndromes.[7–11] With respect to the kidney, the kidney fluid balance system[118] and the systemic renin-angiotensin-aldosterone system (RAAS)[119] have also been extensively studied and established as BP control systems, with several genes identified as anti-hypertensive targets.[41] Additionally, the intra-renal RAAS has been studied and implicated in hypertension.[120] Thus, we must consider both classes of genes/loci to understand the mechanisms of BP variation, and determining their contributions to BP physiology. Furthermore, understanding the potential involvement of non-renal tissues and connecting tissue-specific information to blood pressure should also reveal greater specificity about the entire gamut of mechanisms underlying its regulation.

In order to identify the involvement of specific genes and their networks, and to begin to dissect additional mechanisms underlying BP regulation, we need to identify the

tissues whose biology are essential to this trait. While specific BP regulation mechanisms centrally involving the kidney have been established, the various BP control systems, including the systemic RAAS, are known to involve or thought to affect numerous other tissues, including, but not limited to, the adrenal glands, brain, heart, and vasculature.[121] Thus, there is a need to identify non-renal tissues of interest for understanding BP control. At this time, the vast majority of BP GWAS loci remain to be characterized in terms of specific genes and variants of interest and in terms of their functional context, and, thus, their tissue effects are unknown. This can be performed through the evaluation of enrichment of GWAS loci among expression quantitative trait loci (eQTLs), as has been done in previous studies.[122–124] However, these analyses have so far been limited to a few tissue sources. Consequently, their roles across diverse tissues is unknown, an aspect critical to identifying tissues of relevance for a given phenotype when genetic pleiotropy is widespread.

The development and expansion of public genomic resources have introduced multidimensional functional analyses into standard genetic analyses. The Encyclopedia of DNA Elements (ENCODE) Project (https://www.encodeproject.org/) has generated open chromatin, RNA and DNA sequencing, genotyping, and histone modification data, among other data types. The Genotype-Tissue Expression (GTEx) Project (https://www.gtexportal.org/) includes genotype and expression data across 53 tissues and is useful as reference transcriptome and expression quantitative trait locus (eQTL) dataset. As an example, and relevant to the studies described in this chapter, consider that standard variant annotation analysis usually occurs at the level of the genome. However, these public resources have enabled the development of an annotation score, deltaSVM,[48]

in which the quantitative impact of a variant on gene regulation is predicted specific to a tissue or cell type, based on a reference training set of regulatory regions from tissues or cell types of interest.

Given the scientific need, and the availability of these resources, the primary aims of this chapter focused on identifying tissues of interest for BP regulation to functionally identify and contextualize specific causal genes and variants at some subset of the hundreds of known BP GWAS loci. We also aimed to extend studies of the 20 monogenic BP syndrome genes beyond their known coding variation into their regulatory variation in the kidney.

To identify tissues of interest to BP GWAS loci, we performed an eQTL enrichment analysis in the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH) Genetic Epidemiology Research on Adult Health and Aging (GERA)[125,126] study and determined that associated variants at BP GWAS loci are enriched in eQTLs specific to the aorta and tibial arteries. For each artery, we next connected groups of proximal putative regulatory variants within and around each gene to both the gene's expression and also to BP traits, inferring that the gene's expression in a potentially relevant tissue affected the regulation of BP. To accomplish this, within each artery, we identified putative CREs, and by extension, putative CRE variants, for every gene, and tested these variants in aggregate for association with BP in the GERA study, as well as with expression in the GTEx study, in each of the arteries. We used the sequence kernel association test (SKAT)[43] for these association analyses, with each variant weighted by their deltaSVM scores, to up-weight variants with greater predicted effects on gene regulatory activity. We supplemented the expression analyses with use of

the software, MetaXcan[127], to identify whether the predicted expression of genes in each individual could be associated with BP. As all of these BP gene discovery analyses were novel, we first examined genes for a trait for which there is strong functional evidence (a positive control), such as the cardiac trait QT interval in the Atherosclerosis Risk in Communities (ARIC)[60,61] study, and demonstrated the utility of these analyses for identifying genes relevant to the regulation of BP. It is expected that the QT interval is regulated by the heart only. Finally, we examined the effects of putative regulatory variation for the 20 monogenic genes in four available kidney cell types using analyses as described above to test for a group effect on association with BP.

The results of these analyses suggested specific candidate genes at novel and previously replicated blood pressure loci regulating BP variation between individuals, genes that will aid in our understanding of the BP control mechanisms. We expected that these genes will facilitate more comprehensive analyses of gene regulatory networks involved in BP control.

Methods

Study participants and summary of genotypes, phenotypes, and association results from Hoffmann et al. Nat Genet 2017

The full descriptions of the studies, phenotypes, and association results are in Hoffmann et al.[33] and are briefly recapitulated here. The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, part of the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), consists of individuals from five populations; the majority is non-Hispanic white (EUR), with the remainder including

Latino, East Asians, African Americans, and South Asians. A total of 99,785 individuals were analyzed, of which 80,792 were EUR individuals. The populations were each genotyped on custom population-specific Affymetrix Axiom SNP genotyping arrays, and imputed to the 1000 Genomes Phase I Integrated Release Version 3 haplotype panel. Analyses of GERA alone, with the results of the International Consortium for Blood Pressure (ICBP, n=69,396) study, and with the ICBP and the UK Biobank (UKB, n=152,081) study identified 316 novel BP loci. Combined with the set of replicated BP GWAS loci available at that time, there were a total of 390 loci we considered to be of interest in BP regulation. Of these, 367 had minor allele frequency (MAF) > 0.001 in the GERA EUR study, which was used as the reference population for the eQTL analyses described below.

For the purpose of several of the analyses described in this chapter, we used these association results, as well as summary statistics available from 80,792 GERA EUR individuals from the Hoffmann et al. study, and genotypes from a subset of 71,404 GERA EUR 'unrelated' individuals (third degree or beyond, pruned by the KING software for relationship inference.)[71] We converted genotypes prepared in the Hoffmann et al. study after imputation from IMPUTE2 genotype probabilities format to PLINK 'hard' calls, setting genotypes with uncertainty greater than 0.25 to missing, and retaining variants with < 10% missing data, a Hardy Weinberg equilibrium test p-value < $1x10^{-6}$, and imputation quality score ≥ 0.3. In order to report single-variant summary statistics within this subset of individuals, we used the --assoc option for analysis of a quantitative trait (Wald test) with PLINK v1.9.[69]

We analyzed adjusted systolic (SBP) and diastolic (DBP) blood pressure in this study, as also described in Hoffmann et al.[33]

ARIC genotypes, phenotypes, and association

The Atherosclerosis Risk in Communities (ARIC) cohort is a longitudinal population-based study of 15,792 individuals, comprised of 11,478 European-Americans (EUR) and 4,266 African-Americans (AA) from four study centers: Washington County, MD; Forsyth County, NC; Jackson, MI; and, Minneapolis, MN.[60,61] The initial examination occurred from 1987-1989, with participants aged between 45 and 64 years. Subsequent examinations occurred in 1990-1992, 1993-1995, 1996-1998, and 2011-2013, with the most recent visit (6) beginning in 2016. We analyzed 9,083 individuals of European ancestry with genotypes and QT interval phenotypes available in this study at baseline. The genotyping of these samples on the Affymetrix genome-wide Human SNP Array 6.0, quality control, and imputation to the 1000 Genomes Phase I Integrated Release Version 3 haplotype panel are described elsewhere.[128,129] We converted IMPUTE2 genotype probabilities to PLINK 'hard' calls, setting genotypes with uncertainty greater than 0.25 to missing, and retaining variants with < 10% missing data, a Hardy Weinberg equilibrium test p-value < $1x10^{-6}$, and imputation quality score ≥ 0.3, (as was done for the GERA study; see above). The phenotypes were analyzed as previously described[130] with QT residuals generated by adjusting raw QT intervals for resting heart rate, age, and sex. Summary statistics were generated for single variants using the --assoc option for analysis of a quantitative trait (Wald test) with PLINK v1.9.[69]

GTEx genotypes and expression

We analyzed genotypes and expression data from the Genotype-Tissue Expression (GTEx) Project (https://www.gtexportal.org/) v6p for the SKAT analysis (see below) from the aorta, tibial artery, heart left ventricle, and heart atrial appendage tissues. Normalized expression was analyzed for these tissues, with the top three principal components, available PEER factors (15-35, depending on sample size), genotyping array platform, and sex used as covariates, all available from the GTEx portal. We used SNP-gene associations from the associated * .v6p.all_snpgene_pairs.txt.gz files from the authors' eQTL analyses.

eQTL enrichment and DAVID analyses: Nat Genet. 2017[33]

We conducted an eQTL enrichment analysis and an annotation analysis using the GERA EUR data as the reference panel for locus definitions based on linkage disequilibrium (LD).

The following has been reprinted from the accepted preprint of Hoffmann et al. Nat Genet. 2017[33]

eQTL enrichment analysis

To carry out tissue-specific eQTL enrichment analysis, we used 44 tissue types with at least 70 samples available from the GTex project[131] in addition to seven kidney eQTLs.[132] We used 367 sentinel variants from previously reported SNPs and the three discovery stages presented here available in the GERA cohort with MAF>0.001. Next, 100 sets of 367 random pseudo-sentinel variants were selected matching the MAF to the

original 367 (within ±0.5%). Within each set, the selection was done without replacement; the match for each variant was selected one at a time, and selection of the subsequent variant excluded all previously selected variants, as well as all variants within ±0.5 Mb of all previously selected variants.

Enrichment was tested at both the sentinel SNP level and locus level, conceptually similar to.[122] At the sentinel SNP level, the number of the 367 variants that were also eQTLs in any of the 45 tissues was counted. At the locus level, variants in high LD (r^2 >0.8) with any of the 367 variants were examined for overlap with eQTLs, and if at least one variant within the locus was also an eQTL, the locus was counted. Subsequently, this was repeated for each of the 100 randomly generated sets to observe if an enrichment of eQTLs was visible in the GWAS set. In order to assess which of the 45 tissues were driving the enrichment, counts were also computed per tissue. For each tissue, an upper-tailed p-value for enrichment of the GWAS count was calculated with a Z-score computed using the mean and standard deviation of the null distribution for that tissue.

DAVID analysis

Annotation of genes surrounding the sentinel variants was conducted with DAVID 6.8 beta (non-beta was 6 years old).[133,134] Genes within a ±0.5Mb window of each of the 390 sentinel variants were selected, as defined by GENCODE v19 GTF.[135] Subsequently, those with at least one significant eQTL in tissues identified from the previous enrichment analysis were included in the final list for analysis. Functional annotation analysis was run on the *Homo sapiens* background with default annotations in the categories of disease, functional categories, gene ontology, pathways, and protein

domains, as well as with default parameters, retaining terms with at least two assigned genes. Annotation terms meeting Benjamini-Hochberg $P<0.05$ (adjusting for the number of terms) were considered significant.

End of reprint

Partitioned heritability analyses

We used the stratified LD score regression method and software[136] for estimating the heritability of the trait partitioned by genomic element using summary statistics for SBP and DBP from 80,792 GERA EUR individuals. The mungestats.py script was used to format the summary statistics as appropriate, and the baseline model with 53 categories which include coding, UTR, and intronic regions, in addition to various open chromatin and histone modification annotations as described by the authors, as well as the 1000 Genomes Phase 3 reference files with the weights from their weights_hm3_no_hla.tgz file, which were provided and described by the authors on their website (https://github.com/bulik/ldsc/wiki/Partitioned-Heritability), were used to conduct the analyses.

Construction of putative regulatory element maps

Much of the statistical analyses described here required identification of putative enhancers that were active in tissues of interest. Putative regulatory element maps for the arteries were constructed using datasets available from ENCODE considered as representative of the aorta, tibial artery, and certain parts of the kidney (downloaded May 23, 2017). Open chromatin regions that were detected by either DNAse-seq or ATAC-seq

were taken as putative regulatory elements; for the arteries, enhancer-associated H3K27ac or H3K4me1 histone modifications were used to subset those open chromatin regions that overlapped and were taken as putative enhancers. We completed this work as an extension of previous work in the generation of a cardiac CRE map (Lee et al., unpublished).

As there were no whole aorta open chromatin datasets, we merged five artery open chromatin datasets for this tissue, available among the adult tissue datasets from the ENCODE project. These included DNase-seq datasets from two aortic cell types (smooth muscle (ENCSR000EIH) and adventitia (ENCSR000EMC)) and two pulmonary artery cell types (endothelial cell (ENCSR000EOG) and fibroblast (ENCSR000EOH)), as well as one ATAC-seq dataset from whole tibial artery (ENCSR630REB). The tibial artery dataset was available in raw FASTQ format at the time of download, and was processed using the Kundaje lab protocol (https://github.com/kundajelab/atac_dnase_pipelines), with the exception of the BAM processing steps, including sorting, cleaning, duplicate marking and alignment filtering, retaining reads with quality score >30 and if properly paired, with samtools[137] v1.3 and Picard Tools[138] v2.9.2, which were completed as described in (https://github.com/tobiasrausch/ATACseq/blob/master/src/atac.sh). We used five whole, ascending, or thoracic aorta H3K27ac datasets (ENCSR069UMW, ENCSR015GFK, ENCSR318HUC, ENCSR519CFV, and ENCSR322TJD).

For tibial artery, we used the tibial artery ATAC-seq dataset that was also used as a subset of the open chromatin regions for aorta above (ENCSR630REB), as well as one H3K4me1 dataset available for whole tibial artery (ENCSR233LCT). The peak calling for the H3K4me1 dataset was redone from the available filtered BAM file with MACS2

with the --broad option to call broad peaks, using the parameters q=0.05 for the narrow peak cutoff, and q=0.2 for the broad peak cutoff, with extension size 400 bases.

While no whole kidney open chromatin data were available, there were four cell types available from the ENCODE project from adult/non-fetal data: renal cortical epithelial cell (ENCSR000EOK), glomerular endothelial cell (ENCSR000EOM), epithelial cell of the proximal tubule (ENCSR000EPW), and glomerular visceral epithelial cell from a 3-year-old child (ENCSR785BDQ). No histone modification data were used, so open chromatin regions were considered as putative regulatory regions as a whole category.

For all experiments, where replicates were available, the one with the fewer peaks called was used as the reference replicate with peaks overlapping the replicate within each experiment retained. The artery open chromatin peaks were merged with the bedtools v2.26 merge utility.[139] To prepare the open chromatin regions for each tissue for training with gkm-SVM, we selected 600bp regions to represent each putative regulatory element. Peaks shorter than 600bp were extended 300bp from the midpoint, and peaks longer were subset to the 600bp region with the maximum signal over all combined datasets for each tissue. The five H3K27ac datasets for aorta were merged using bedtools merge, and the peaks from this merged set and the tibial artery H3K4me1 dataset were extended by 100 bases in both directions, as we considered the boundaries of histone modification peaks to be inexact.[140,141] We retained those open chromatin regions overlapping a histone modification peak (after the 100bp extension) by at least 50% for training for the aorta and tibial arteries. Nearly all the datasets were available on genome

build hg38, so these analyses were carried out on this build and then lifted over to hg19 for compatibility with downstream analyses.

Generation of deltaSVM scores

Each set of peaks (hg19 build) for the six tissues or cell types (aorta, tibial artery, and each of the four kidney cell types) was used as a positive training set to build gkm-SVM models,[142] as described in other work[142–144] with some modifications. The LS-GKM[145] software with the default parameter set was used for training. Regions from chromosome 9 were excluded from the training and used for evaluating the adequacy of the trained models. The final models were generated by averaging gkm-SVMs from 10 independent negative training sets for each of the tissues or cell types. Subsequently, for each tissue or cell type, a deltaSVM score was calculated for each of 10,041,372 variants from the 1000 Genomes European-ancestry population, defined as the difference in gkm-SVM scores for reference and alternate alleles for that variant, as previously described.[48]

Gene-based testing with SKAT

We used the sequence-kernel association test (SKAT)[43] within the SKAT R package[146] to test genes with median RPKM ≥ 0.3 in GTEx samples for the aorta (n=197) and tibial (n=285) arteries with their respective variant sets. For each gene, we tested all variants within 50Kb of the gene start or end, inclusive of the entire gene body, per GENCODE v19 annotations (https://www.gencodegenes.org/releases/19.html). The weights used were taken as the absolute value of the deltaSVM score for each variant to reflect its predicted impact; for comparison, we also ran SKAT using default weights

with beta density parameters (weights.beta=c(1,25), which up-weights rare variants as compared to common variants), as well as equal weights to all variants (weights.beta=c(1,1)). We tested association of each gene with adjusted SBP and DBP phenotype residuals (see above), as well as the GTEx normalized expression data with covariates (release v6p, https://www.gtexportal.org/), from the aorta and tibial arteries. We restricted our primary analyses in each of the kidney cell types to the 20 monogenic genes. We additionally tested tissue- or cell-type-specific groupings in the ARIC dataset with the adjusted QT interval phenotype using the sets for the heart and heart tissues from GTEx, arteries and kidney cell types, as described above.

Prediction of gene expression association with blood pressure

We used the MetaXcan[127] software with prebuilt HapMap training models for the GTEx (https://www.gtexportal.org/) tissues aorta and tibial arteries, provided by the authors at http://predictdb.hakyimlab.org/, with summary statistics from 80,792 GERA EUR individuals for SBP and DBP. We also used the software with the provided models for heart left ventricle and atrial appendage, for the QT interval phenotype analysis using summary statistics from 9,083 ARIC EUR individuals. MetaXcan is an extension of the PrediXcan[147] method, which predicts gene expression from genotypes and tests association of predicted expression with phenotypes using summary association results.

Statistical significance

Significance was determined for analysis using the Bonferroni method for multiple test correction to adjust for the number of genes within each analysis. We made

no additional adjustment for the number of tissues, in part due to the correlation of specific subsets (the arteries, and individual kidney cell types), and as we examined genes across multiple analyses, for phenotype and for expression, to note their significance.

Results

We conducted several tissue-specific analyses to identify tissues and genes of interest for BP regulation in the GERA study. We initially focused on identifying tissues relevant to BP GWAS loci, and subsequently expanded on this by using tissue-specific information to analyze putative cis-regulatory variation of genes in these tissues to identify specific genes and variants of interest at these loci. We also studied putative regulatory variation of 20 monogenic genes in several kidney cell types. These analyses are described below.

The following has been reprinted from the accepted preprint of Hoffmann et al. Nat Genet. 2017[33]

Expression Quantitative Trait Loci analysis in different tissues and enrichment analysis for functional elements

We next investigated whether the previously identified and all newly identified loci (367 sentinel variants with MAF>0.001), co-localized with Expression Quantitative Trait Loci (eQTLs) in certain tissues. We used eQTLs from 44 Genotype-Tissue Expression (GTEx) tissues and kidney eQTLs.[131,132] Across all tissues, there are 186 significant eQTLs at the sentinel level, and 213 at the locus level. We determined for each tissue whether the number of eQTLs observed (either by sentinel SNP or by locus)

was greater than expected by chance, where expectation was derived from a random sampling of SNPs and loci (see Online Methods). We ranked the tissues by eQTL *P* value, both for the sentinel SNP and locus analysis. As tissues with higher numbers of eQTLs are generally expected to overlap SNP sets to a greater degree, it may be expected that enrichment is greater in the tissues with the highest number of eQTLs simply because of chance overlap with the GWAS set, especially when eQTLs in tissues relevant to the phenotype are also found in these tissues. To observe whether the enrichment visible for a given tissue is greater than what we expect relative to the total number of eQTLs it contains, we examined the relationship between *P* value and total eQTL count per tissue (Figure 1). We observed that the aorta and tibial artery are clear outliers as compared to the other tissues, even accounting for total number of eQTLs in those tissues.

Enrichment analysis for functional elements

We subsequently investigated whether the genes near the sentinel variants are enriched for certain functional pathways. We included genes within ±0.5Mb of the 390 sentinel variants with a significant eQTL in either of the two tissues identified above (aorta and tibial artery). We identified 2,013 genes near all 390 sentinel variants (see **Online Methods**) and tested for enrichment of functional annotations. Using DAVID 6.8,[133,134] 1,480 had annotations available, producing 26 significant annotation terms (Benjamini-Hochberg $P<0.05$, Table 1), without a clear functional pathway emerging.

End of reprint

Partitioned heritability of BP

We examined BP heritability for SBP and DBP for 80,792 GERA EUR subjects with stratified LD score regression (LDSC) across several functional categories[136] to identify functional categories in which BP heritability was enriched, and found that the top-ranked categories of enrichment were those of enhancer-associated histone marks H3K27ac, H3K4me1, and the Hnisz "super-enhancer" category (Tables 2-3). This is in accordance with a previous study in which BP heritability was determined to be mostly from within DNaseI hypersensitivity sites,[6] and, taken together with the results of the eQTL enrichment analyses, supports the study of regulatory elements in specific tissues of interest for BP.

Tissue-specific identification of candidate genes

With knowledge of tissues highly relevant to characterizing BP GWAS loci, our next aim was to test each gene's putative cis-regulatory variation for association with both blood pressure and expression, in a tissue-specific context. This is expected to assist in identifying novel genes of interest, as well as provide tissue- or cell-type-specific information about known genes.

We first constructed putative regulatory element maps for the aorta and tibial arteries, as well as four kidney cell types (renal cortical epithelial cell, glomerular endothelial cell, epithelial cell of proximal tubule, and glomerular visceral epithelial cell), because of the involvement of the kidney in blood pressure regulation, using data from the ENCODE project (Table 4). These were completed as an extension of the construction of a cardiac putative regulatory element map in our previous work (Lee et

al., unpublished). We specifically focused on identifying putative enhancers for the aorta and tibial arteries (see Methods). We subsequently used these maps for training with the software gkm-SVM, in order to generate deltaSVM functional scores for all variants from the 1000 Genomes European population, to be tested for association on a gene-level basis. The performance for each model was as follows: aorta, AUC=0.88 trained on 61,083 peaks; tibial artery, AUC=0.84 trained on 117,994 peaks, renal cortical epithelial cell, AUC=0.97 trained on 108,858 peaks; glomerular endothelial cell, AUC=0.96 trained on 55,191 peaks; epithelial cell of proximal tubule, AUC=0.96 trained on 76,889 peaks; and glomerular visceral epithelial cell, AUC=0.96 trained on 70,934 peaks. A possible reason for the improved performance of the renal cell types is that open chromatin regions from individual cell types were used, which may have been optimal for training, whereas the mixture of cell types comprising the arteries may have affected the performance due to biological complexity. The magnitude of the deltaSVM score for a variant reflects its predicted impact on regulatory functional activity, while its sign reflects the prediction with respect to the reference or alternate alleles. Therefore, to represent the predicted impact of each variant irrespective of allele, we show the distributions of the absolute values of the deltaSVM scores for the arteries and kidney cell types in Figure 2.

As our emphasis in this section is to connect a gene's putative CRE variants to both a phenotype of interest and to its expression in relevant tissues (see Figure 3 for a schematic describing the BP association analyses), we first describe the overall analysis scheme as applied to a general phenotype of interest. We then describe how we applied these analyses, first to the QT interval in the ARIC study, as proof of principle to

demonstrate the utility of these analyses, and then to our BP traits of interest in the GERA study.

We defined a gene's "cis"-regulatory variants in this analysis as those variants falling in putative CREs within 50Kb of the gene's start and end, and tested the aggregate effect of these CRE variants for each gene using SKAT,[43] for association with the phenotype(s) of interest in the relevant population, which were SBP and DBP in the GERA study, and QT interval in the ARIC study. SKAT is a test that has generally been used to study groups of variants together and is useful when variants can have bidirectional effects; rare variants are more highly weighted than common variants by default. In addition to the default weights, we ran the analysis using equal weights for all variants. We finally used the tissue- or cell-type-specific deltaSVM scores for the analyzed variants as weights for a custom SKAT test; the greater an effect the variant is predicted to have on functional regulatory activity, the higher the score.

We then tested these groupings with expression data from GTEx v6p in the tissues of interest to link variants in the genes of interest to their expression in the relevant tissues. The groupings tested in the GTEx data with expression were not always identical to the groupings tested in the GERA or ARIC studies because of differences in imputation quality score filtering, missingness of genotypes from genotype probabilities to hard call conversion, and variants present in the reference populations studied. However, this analysis still connects a given gene to its expression and to the phenotype via a highly overlapping set of putative proximal regulatory variants, and was completed this way to test the most complete set of variants available meeting our criteria, in each of the datasets. In addition to testing putative regulatory variants with gene expression in

GTEx, we used the recently developed MetaXcan[127] software to augment the SKAT results and to identify any new associations by this method. This software predicts the association of gene expression with the phenotype given genotypes for the population of interest based on training from reference genotypes and expression data.

As mentioned earlier, we considered the cardiac trait QT interval first to demonstrate proof of concept for tissue-specific gene identification. The QT interval is the measurement between the Q and T waves in the electrical conduction system in the heart[148] with about ~30% heritability.[149–152] Two of the genes with major effects as estimated through a GWAS that have also been implicated and functionally validated in the heritability of QT interval are *NOS1AP*[152–154] and *SCN5A*,[154,155] and so are the focus for this particular analysis. We analyzed the association of the QT interval adjusted for heart rate, age, and sex at baseline in 9,083 ARIC EUR individuals with putative CRE variants in the *SCN5A* and *NOS1AP* genes using SKAT and MetaXcan; these results are depicted in Tables 5-6. The tissues available from GTEx that we considered to be of relevance for QT interval for expression analyses are the heart left ventricle and atrial appendage tissues; for comparison, we also show the six tissues or cell types we analyzed in this study for blood pressure in these tables. A heart-specific effect is visible for *SCN5A* (deltaSVM p, p.SKAT.qt.dsvm =4.86 x 10^{-4}; equal weighting p, p.SKAT.qt.eq=1.65 x 10^{-4}) and not for the arteries or kidney cell types (Table 5); a much larger number of variants were analyzed for heart as there are more open chromatin regions within and immediately surrounding the gene in the heart compared to other tissues. The association with expression is not visible here, however. The *NOS1AP* gene shows a strong signal across all tissue/cell types (Table 6), indicating the presence of

individual variants driving the signal in open chromatin regions across all of the analyzed tissue/cell types. However, it is interesting to note that the SKAT analysis in GTEx heart left ventricle tissue shows the greatest evidence for association (p.SKAT.GTEx.dsvm=3.85 x 10^{-4}; p.SKAT.GTEx.eq=1.51 x 10^{-4}), with the next strongest signal from tibial artery (p.SKAT.GTEx.dsvm=2.22 x 10^{-3}; p.SKAT.GTEx.eq=0.012). The MetaXcan results were uninformative in this case; there were no models available for *NOS1AP* in either heart tissue or for *SCN5A* in heart left ventricle (p.MetaXcan=0.011 in heart atrial appendage).

Considering both sets of effects, certainly variants with detectable signals present in open chromatin regions specific to the relevant tissue/cell types will allow the detection of a tissue-specific signal, as for *SCN5A*. It also appears, however, that gene-level signals may be captured by analyses in which all variants are weighted equally, and when local open chromatin boundaries across tissues/cell types overlap considerably, especially when variants with strong signals are present within these shared regions, we will not necessarily be able to differentiate between different tissue/cell types. Weighting with the tissue-specific deltaSVM scores introduces an additional tier of tissue specificity and is based on global open chromatin differences. While we see some of this difference for *SCN5A* with the ENCSR000EOM results (p.SKAT.qt.dsvm=0.077; p.SKAT.qt.eq=0.0013), it is for *NOS1AP* in particular that we can see that the aorta and ENCSR000EOK groupings are down-weighted by ~10 orders of magnitude using deltaSVM weights compared to using equal weights. Finally, using the default weights shows least concordance with the other two sets of results, indicating that for this analysis, rare variants are not driving the signal as compared to common variants.

We then applied these analyses to our tissues of interest for BP regulation, which included the aorta, tibial artery, and the four kidney cell types, in a subset of 71,404 unrelated GERA EUR individuals. We tested 14,035 genes expressed at RPKM ≥ 0.3 in 197 aorta GTEx samples and 13,920 genes expressed at RPKM ≥ 0.3 in 285 tibial artery GTEx samples for the SKAT analyses. We used summary statistics already available from 80,792 individuals[33] to maximize the sample size for which the MetaXcan analyses were run, for the aorta and tibial arteries. Results for each of the arteries are presented in Tables 7-10. In some cases, shared variants drive the positive signal for multiple genes at the same locus; expression in the relevant tissue or cell type may pinpoint a specific gene. However, it may be noted that the genes *CERS5*, *COX14*, and *RP4-605O3.4* are all present at the same locus in the arteries (Tables 7 and 9), but evidence of expression association is present for multiple of these genes; this may be indicative of proximal variants affecting the different genes, or pleiotropy of single variants affecting the expression of multiple genes.

While there are several genes in each analysis with interesting associations with BP traits, here we highlight six genes that are significant by deltaSVM-weighted SKAT in either of the arteries and also have strong expression support, reaching significance either from MetaXcan or the analogous GTEx SKAT analysis (using the same deltaSVM p-value Bonferroni-determined cutoff as in the phenotype SKAT analysis, with the majority of variants tested in the phenotype SKAT analysis also tested in the expression SKAT analysis). These include: *NOV* (tibial artery, DBP); *ULK4* (aorta and tibial arteries, DBP); *SDCCAG8* (aorta, DBP), and *CLCN6* (tibial artery, SBP), reaching significance in the relevant MetaXcan analyses; additionally, though not significant by

MetaXcan but by SKAT in GTEx, are the two genes *MTHFR* (tibial artery, SBP) and *C10orf32* (tibial artery, SBP). As previously mentioned, though variant sets tested for each gene for association with phenotype and tissue-specific expression were not always identical, we examined which variants were driving the signal, as displayed in Tables 11-12 (aorta and tibial arteries, respectively). Most of these genes (*NOV*[16], *ULK4*[26], *CLCN6*[100], *MTHFR*[13,28], *C10orf32*[28]) are present at or near previously replicated BP GWAS loci; *SDCCAG8* was identified as part of Hoffmann et al.[33] It is noteworthy that both *CLCN6* and *MTHFR* are neighboring genes, but have independent expression support in the same tissue.

We also studied the genes involved in monogenic forms of hypotension or hypertension in four kidney cell types available from the ENCODE project: renal cortical epithelial cell (ENCSR000EOK), glomerular endothelial cell (ENCSR000EOM), epithelial cell of proximal tubule (ENCSR000EPW), and glomerular visceral epithelial cell (ENCSR785BDQ). As the expression data available for kidney are insufficient, we studied each cell type individually and carried out only SKAT analyses for these genes; the results are in Table 13-14. The most notable result is that of *CYP17A1*, which shows an effect ($p \sim 10^{-5} - 10^{-7}$) across all four cell types in the un-weighted variants analyses for SBP only, and more specifically, only in the glomerular endothelial cell (ENCSR000EOM) ($p.SKAT.dsvm.ENCSR000EOM = 1.26 \times 10^{-7}$) in the deltaSVM-weighted results. However, as *C10orf32* is a gene of interest at the same locus based on the tibial artery results above, we examined and noted that the results are somewhat similar for this gene, though not as striking, due to variant set sharing in the SKAT analyses for these genes (deltaSVM p-values: ENCSR000EOK, 8 variants, $p = 1.33 \times 10^{-3}$;

ENCSR000EOM, 12 variants, $p=3.87 \times 10^{-5}$; ENCSR000EPW, 10 variants, $p=8.25 \times 10^{-4}$; ENCSR786BDQ, 9 variants, p=0.027). The breakdown of individual variants analyzed for these two genes is in Table 15. The variant rs3824754, with an SBP association $p=1.40 \times 10^{-11}$, appears in the groupings of both genes for all four cell types, but has the highest deltaSVM magnitude in the endothelial cell. Additionally, there is a set of four variants with SBP association $p<1 \times 10^{-4}$ (rs284853, rs284854, rs284855, rs284856) which only appear in the ENCSR000EOM groupings. We observed that while *CYP17A1* was similarly associated with, or demonstrated evidence of association with, SBP in the deltaSVM and unweighted variants analysis (aorta deltaSVM $p=1.12 \times 10^{-5}$, 34 variants; tibial artery deltaSVM $p=6.78 \times 10^{-8}$, 15 variants, Table 9), the analysis of variants in GTEx (33 variants for aorta and 14 variants for tibial artery) did not reflect any significant association (p>0.01). Contrastingly, *C10orf32* demonstrated association of evidence with SBP (aorta deltaSVM $p=1.26 \times 10^{-5}$, 17 variants; tibial artery deltaSVM $p=6.73 \times 10^{-8}$, 8 variants, Table 9) and with expression in GTEx (aorta deltaSVM $p=1.22 \times 10^{-13}$, 15 variants; tibial artery deltaSVM $p=1.87 \times 10^{-14}$, 7 variants, Table 9). The same four variants unique to the ENCSR000EOM groupings above with strong associations with SBP are also present in the artery groupings. As is evident from Tables 11 and 12, three of these variants (rs284854, rs284855, rs284856) are eQTLs for *C10orf32* in the aorta and tibial arteries; these variants do not show association with *CYP17A1* expression in these tissues (all p>0.03 for aorta, all p>0.21 for tibial artery, from eQTL data available from the GTEx portal (https://www.gtexportal.org/)). This may reflect an endothelial-cell-specific effect for C10orf32 rather than a tissue-type effect, especially as this locus has been identified in several previous BP GWAS studies,[26,28,156,157]; it may

also not be very informative for the kidney, though suitable expression data for kidney would be required to assess this.

Discussion

Our combined analyses identify two specific potential tissues – the aorta and tibial arteries – relevant to blood pressure regulation. We have also subsequently identified several genes with regulatory variants linking significantly to BP traits and to expression data, most at previously replicated BP loci. Although the involvement of the kidney is well established in BP regulation, given the systemic nature of hypertension, we sought to identify genes at any of the hundreds of BP GWAS loci in a tissue-specific manner, by taking advantage of the additional information provided by tissue-specificity. We examined groupings of multiple proximal and putatively causal variants defined around genes within a single tissue in order to identify specific genes of interest.

We identified six genes of interest using artery-specific information: *NOV*, *ULK4*, *SDCCAG8*, *C10orf32*, *CLCN6*, and *MTHFR*. In addition to its role in the progression of various cancers, the *NOV* gene has been identified as a player in angiogenesis[158,159] and vascular homeostasis.[160] The *ULK4* gene has been previously associated with DBP,[26] and variation in this gene has also been associated with aortic disease and acute aortic dissections.[161] Associations of rare variants in the gene *CLCN6*, encoding a chloride channel, have been identified with BP.[100] Additionally, the association of a homozygous variant (C677T) in its neighboring gene, *MTHFR*, has also been associated with BP and vascular disease[162–165]; more generally, this locus has been identified in large BP GWAS.[13,28] The locus including *C10orf32* has been identified previously[28] and neighbors

the well-studied gene *CYP17A1*. Though we initially examined only the latter among kidney cell types because of its known role in monogenic hypertension, we note that both genes show BP association in endothelial contexts, but it is *C10orf32* that has strong expression support in the arteries in our study, while *CYP17A1* does not, and remains to be characterized.[166] Finally, the gene *SDCCAG8* is a centrosomal protein linked with nephronophthisis-related ciliopathies (OMIM: Senior-Loken Syndrome 7, 613615; Bardet-Biedl syndrome-16, 615993, and Airik et al.[167]), and is expressed in the kidney and lung epithelia.[167]

The MetaXcan software has supported most of these genes highlighted here and identified novel associations, although there were some limitations with the availability of the models for all genes. Additionally, our results indicated that deltaSVM weighting might be validly discriminatory between cell types; this is most evident with the QT interval genes *NOS1AP* and *SCN5A*, and is suggestive of cell-type specificity with the results for *CYP17A1* in the kidney cell types. It may be informative moving forward to characterize these genes at the individual cell-type level in the arteries as well.

Our attempts to expand findings with respect to the 20 genes involved in monogenic forms of hypertension or hypotension were inconclusive. We attribute this to the dearth of public data available for the kidney at this time, and expect that the future availability of such data will resolve some of the issues in further studies.

Acknowledgements and Funding

Thanks to Dr. Dongwon Lee for running the gkm-SVM software to produce the deltaSVM scores used in these analyses.

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

The following acknowledgements for the GERA, ICBP, and UKB studies have been reprinted from Hoffmann et al. Nat Genet. 2017[33]

"We are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment, and Health. Support for participant enrollment, survey completion, and biospecimen collection for the RPGEH was provided by the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and Kaiser Permanente Community Benefit Programs. Genotyping of the GERA cohort was funded by a grant from the National Institute on Aging, National Institute of Mental Health, and the National Institute of Health Common Fund (RC2 AG036607 to CAS and NJR). This research has been conducted using the UK Biobank Resource. This research has also been conducted using access-controlled ICBP data from dbGaP. We thank our

colleagues for making these data available. Data analyses were facilitated by NHLBI grant R01 HL128782 (to AC and NJR). GE receives support from Geneva University Hospitals and The Foundation of Medical Researchers, Geneva."

End of reprint

Chapter 4 Figures

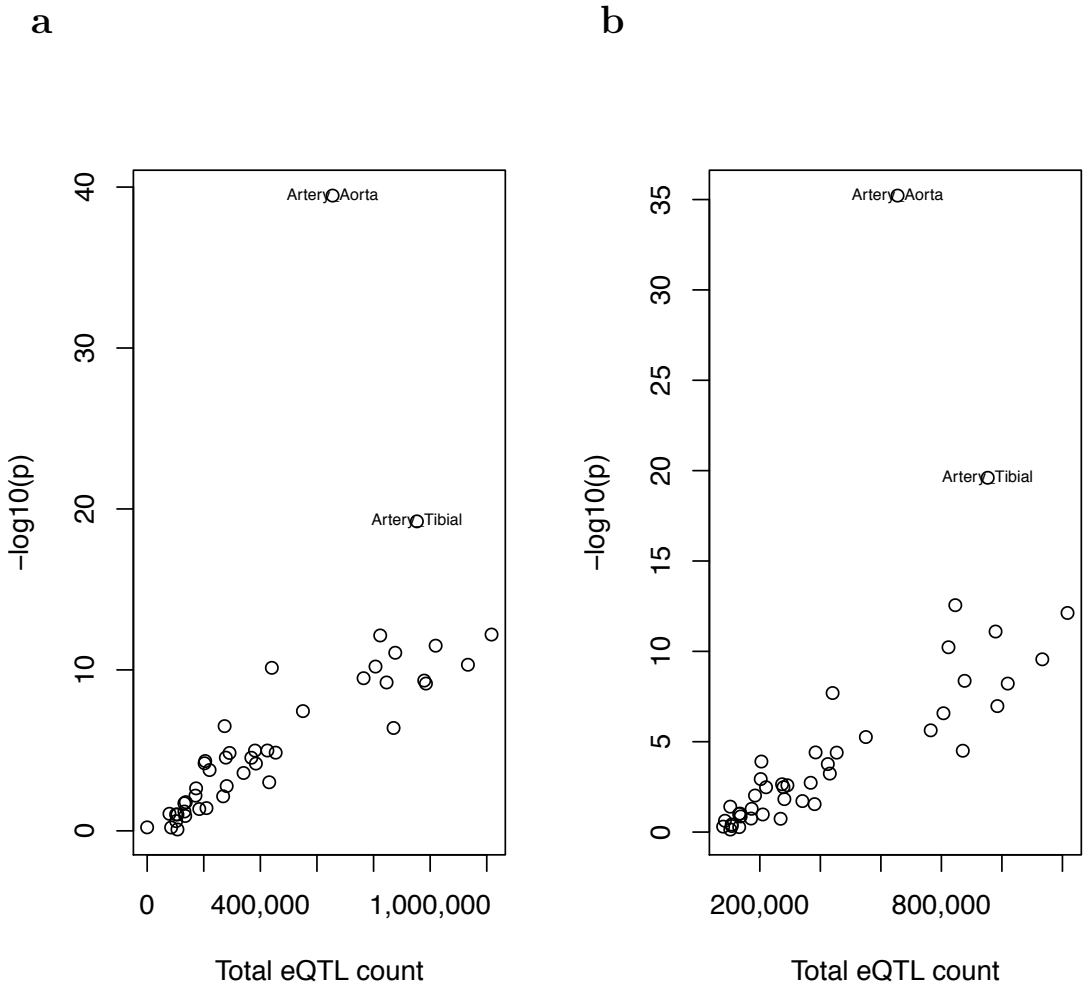

Figure 1. Tissue specific expression quantitative trait loci (eQTL) analysis of 45 tissues. The two outlier tissues, accounting for total eQTL count, are labeled. Tissue total eQTL counts vs P-values at (a) the locus and (b) the sentinel variant.

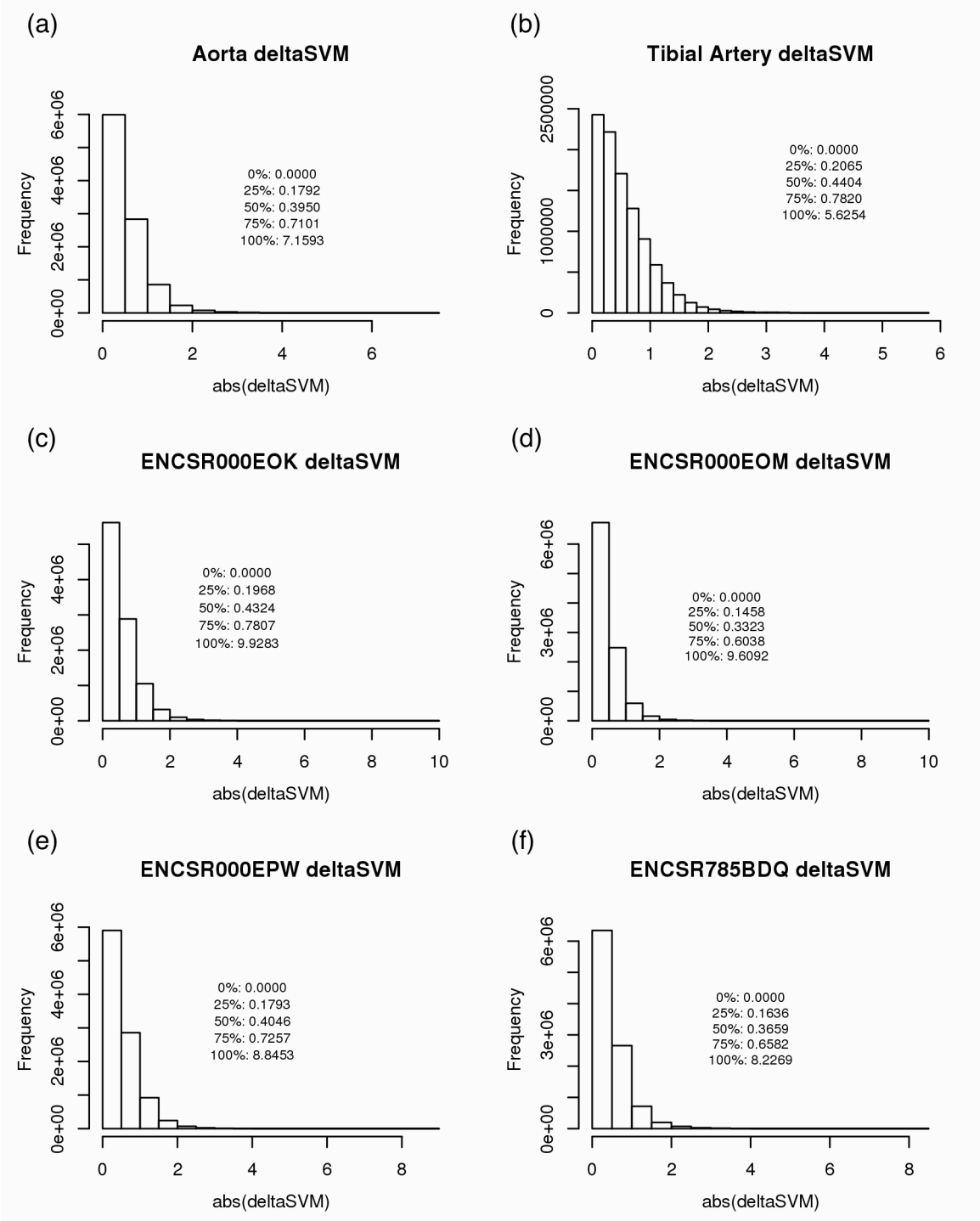

Figure 2. Distributions of absolute value of deltaSVM scores in arteries and kidney cell types. 0%, minimum; 25%, first quartile; 50%, median; 75%, third quartile; 100%, maximum.

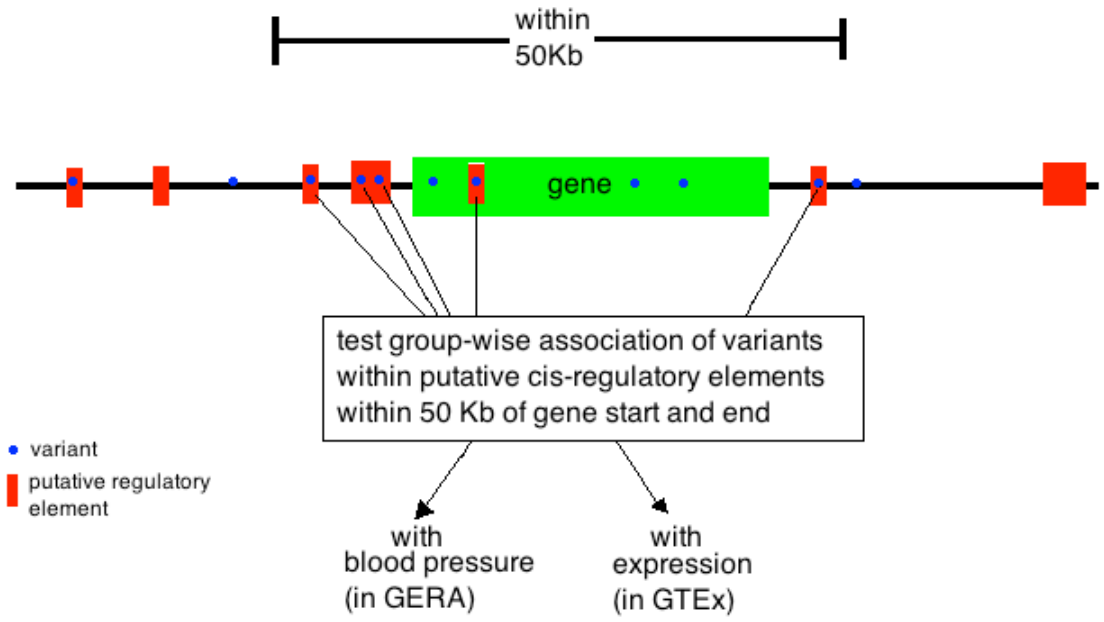

Figure 3. Schematic of BP association analyses in this study. We test association of groups of variants present within open chromatin regions within 50Kb of the gene start or end with both the gene's measured expression in the tissues of interest in GTEx, and with blood pressure in GERA.

Chapter 4 Tables

Table 1. DAVID enrichment analysis.

Category	Term	Count	%	List Total	Pop Hits	Pop Total	Fold Enrichment	Benjamini
UP_KEYWORDS	Alternative splicing	846	57.2	1328	10581	20042	1.21	1.50×10^{-14}
UP_SEQ_FEATURE	splice variant	616	41.6	1316	7761	20064	1.21	1.30×10^{-6}
UP_KEYWORDS	Polymorphism	901	60.9	1328	12023	20042	1.13	1.50×10^{-7}
UP_KEYWORDS	Phosphoprotein	595	40.2	1328	7493	20042	1.2	1.20×10^{-6}
UP_KEYWORDS	Acetylation	301	20.3	1328	3426	20042	1.33	6.80×10^{-6}
UP_SEQ_FEATURE	sequence variant	906	61.2	1316	12444	20064	1.11	9.60×10^{-5}
GOTERM_CC_DIRECT	GO:0016020~membrane	181	12.2	1226	1901	17701	1.37	0.0034
GOTERM_CC_DIRECT	GO:0005813~centrosome	55	3.7	1226	416	17701	1.91	0.0018
UP_KEYWORDS	Protein transport	69	4.7	1328	606	20042	1.72	0.0013
GOTERM_MF_DIRECT	GO:0005515~protein binding	676	45.7	1150	8706	16483	1.11	0.017
UP_KEYWORDS	Golgi apparatus	84	5.7	1328	806	20042	1.57	0.0031
UP_KEYWORDS	Transit peptide	61	4.1	1328	533	20042	1.73	0.0027
UP_KEYWORDS	Cell division	47	3.2	1328	378	20042	1.88	0.0029
GOTERM_CC_DIRECT	GO:0005774~vacuolar membrane	7	0.5	1226	12	17701	8.42	0.015
UP_KEYWORDS	Nucleotide-binding	158	10.7	1328	1774	20042	1.34	0.0045
UP_KEYWORDS	Cytoplasm	373	25.2	1328	4750	20042	1.19	0.0042
UP_KEYWORDS	Proteomics identification	204	13.8	1328	2405	20042	1.28	0.0054
KEGG_PATHWAY	hsa04022:cGMP-PKG signaling pathway	26	1.8	487	162	6891	2.27	0.041

UP_KEYWORDS	Cytoskeleton	105	7.1	1328	1112	20042	1.43	0.0082
UP_KEYWORDS	Transport	169	11.4	1328	1965	20042	1.3	0.01
UP_KEYWORDS	Mitochondrion	103	7	1328	1103	20042	1.41	0.013
UP_KEYWORDS	Endoplasmic reticulum	99	6.7	1328	1057	20042	1.41	0.014
UP_KEYWORDS	ATP-binding	124	8.4	1328	1384	20042	1.35	0.013
UP_KEYWORDS	Oxidoreductase	59	4	1328	571	20042	1.56	0.022
UP_KEYWORDS	Coiled coil	215	14.5	1328	2649	20042	1.22	0.023
UP_KEYWORDS	Cell cycle	64	4.3	1328	640	20042	1.51	0.026

%, percent of total genes assigned to term; Count, number of genes in analysis assigned to term; List Total, all genes in analysis with category annotation; Enrichment, ((Count)/(List Total))/((Pop Hits)/(Pop Total))

Table 2. Partitioned heritability results from baseline model for SBP

Category	Prop_SNPs	Prop_h2 (SE)	Enrichment (SE)	Enrichment_p
H3K27ac_Hnisz.extend.500.bedL2_0	0.23	0.79 (0.05)	1.88 (0.12)	3.39×10^{-11}
Conserved_LindbladToh.extend.500.bedL2_0	0.42	0.74 (0.06)	2.23 (0.19)	1.13×10^{-9}
H3K4me1_Trynka.extend.500.bedL2_0	0.39	0.98 (0.05)	1.62 (0.09)	2.00×10^{-9}
H3K4me1_Trynka.bedL2_0	0.72	1.03 (0.10)	2.43 (0.24)	1.05×10^{-8}
SuperEnhancer_Hnisz.extend.500.bedL2_0	0.17	0.35 (0.03)	2.05 (0.17)	2.25×10^{-8}
SuperEnhancer_Hnisz.bedL2_0	0.61	0.36 (0.03)	2.17 (0.20)	2.61×10^{-8}
Repressed_Hoffman.extend.500.bedL2_0	0.33	0.47 (0.05)	0.65 (0.06)	1.64×10^{-7}
H3K9ac_Trynka.extend.500.bedL2_0	0.13	0.55 (0.06)	2.40 (0.28)	1.59×10^{-6}
H3K27ac_Hnisz.bedL2_0	0.03	0.64 (0.05)	1.64 (0.13)	2.60×10^{-6}
Conserved_LindbladToh.bedL2_0	0.26	0.38 (0.08)	14.64 (2.98)	7.71×10^{-6}
H3K4me3_Trynka.extend.500.bedL2_0	0.17	0.56 (0.07)	2.21 (0.26)	1.00×10^{-5}
H3K9ac_Trynka.bedL2_0	0.42	0.49 (0.09)	3.88 (0.68)	4.47×10^{-5}
H3K27ac_PGC2.extend.500.bedL2_0	0.34	0.60 (0.06)	1.80 (0.19)	4.75×10^{-5}
Enhancer_Hoffman.extend.500.bedL2_0	0.09	0.31 (0.06)	3.45 (0.64)	2.80×10^{-4}
Enhancer_Hoffman.bedL2_0	0.11	0.26 (0.07)	6.14 (1.56)	1.45×10^{-3}
UTR_3_UCSC.bedL2_0	0.54	0.10 (0.03)	8.76 (2.60)	2.88×10^{-3}
DHS_peaks_Trynka.bedL2_0	0.17	0.49 (0.13)	4.39 (1.18)	4.55×10^{-3}
FetalDHS_Trynka.extend.500.bedL2_0	0.13	0.56 (0.10)	1.99 (0.34)	4.63×10^{-3}
FetalDHS_Trynka.bedL2_0	0.28	0.39 (0.11)	4.61 (1.29)	5.58×10^{-3}
TFBS_ENCODE.extend.500.bedL2_0	0.04	0.61 (0.10)	1.80 (0.28)	6.16×10^{-3}
DHS_Trynka.extend.500.bedL2_0	0.08	0.76 (0.09)	1.53 (0.19)	6.40×10^{-3}
H3K27ac_PGC2.bedL2_0	0.27	0.48 (0.08)	1.79 (0.31)	0.012
DHS_Trynka.bedL2_0	0.17	0.50 (0.13)	2.99 (0.79)	0.012
DGF_ENCODE.extend.500.bedL2_0	0.4	0.75 (0.09)	1.40 (0.16)	0.015
Intron_UCSC.bedL2_0	0.01	0.47 (0.04)	1.22 (0.09)	0.017
TFBS_ENCODE.bedL2_0	0.34	0.42 (0.13)	3.22 (0.97)	0.024

H3K4me1_peaks_Trynka.bedL2_0	0.04	0.48 (0.14)	2.85 (0.82)	0.025
H3K9ac_peaks_Trynka.bedL2_0	0.04	0.24 (0.09)	6.14 (2.28)	0.027
Transcribed_Hoffman.extend.500.bedL2_0	0.01	0.63 (0.06)	0.82 (0.08)	0.028
Intron_UCSC.extend.500.bedL2_0	0.13	0.45 (0.03)	1.15 (0.08)	0.052
TSS_Hoffman.bedL2_0	0.01	0.09 (0.04)	5.19 (2.38)	0.081
Coding_UCSC.bedL2_0	0.03	0.06 (0.03)	4.26 (2.29)	0.157
PromoterFlanking_Hoffman.extend.500.bedL2_0	0.46	-0.02 (0.04)	-0.67 (1.17)	0.157
PromoterFlanking_Hoffman.bedL2_0	0.06	-0.03 (0.03)	-4.18 (3.88)	0.179
H3K4me3_Trynka.bedL2_0	0.5	0.24 (0.08)	1.78 (0.59)	0.186
H3K4me3_peaks_Trynka.bedL2_0	0.03	0.15 (0.08)	3.56 (1.99)	0.201
Promoter_UCSC.extend.500.bedL2_0	0.14	0.07 (0.03)	1.89 (0.73)	0.228
Coding_UCSC.extend.500.bedL2_0	0.02	0.11 (0.04)	1.73 (0.61)	0.231
UTR_3_UCSC.extend.500.bedL2_0	0.39	0.06 (0.03)	2.22 (1.06)	0.249
Repressed_Hoffman.bedL2_0	0.04	0.34 (0.12)	0.73 (0.26)	0.295
CTCF_Hoffman.extend.500.bedL2_0	0.02	0.01 (0.06)	0.18 (0.85)	0.331
DGF_ENCODE.bedL2_0	0.76	0.26 (0.12)	1.88 (0.92)	0.339
WeakEnhancer_Hoffman.extend.500.bedL2_0	0.02	0.12 (0.07)	1.40 (0.74)	0.590
Enhancer_Andersson.extend.500.bedL2_0	0.07	0.00 (0.03)	0.15 (1.68)	0.614
TSS_Hoffman.extend.500.bedL2_0	0.02	0.05 (0.04)	1.49 (1.06)	0.646
CTCF_Hoffman.bedL2_0	0	0.05 (0.06)	2.11 (2.46)	0.652
WeakEnhancer_Hoffman.bedL2_0	0.09	0.00 (0.05)	0.16 (2.35)	0.720
UTR_5_UCSC.extend.500.bedL2_0	0.03	0.02 (0.03)	0.64 (1.23)	0.771
Transcribed_Hoffman.bedL2_0	0.35	0.32 (0.11)	0.93 (0.31)	0.811
Promoter_UCSC.bedL2_0	0.01	0.04 (0.04)	1.22 (1.32)	0.865
Enhancer_Andersson.bedL2_0	0.03	0.00 (0.03)	0.35 (5.88)	0.911
UTR_5_UCSC.bedL2_0	0.03	0.00 (0.02)	0.64 (3.62)	0.921
baseL2_0	1	1.00 (0.00)	1.00 (0.00)	-

SE, standard error; Prop_SNPs, proportion of SNPs in category; Prop_h2, proportion of heritability attributed to category

Table 3. Partitioned heritability results from baseline model for DBP

Category	Prop_SNPs	Prop_h2 (SE)	Enrichment (SE)	Enrichment_p
H3K9ac_Trynka.extend.500.bedL2_0	0.23	0.73 (0.07)	3.19 (0.29)	1.97×10^{-12}
H3K27ac_Hnisz.extend.500.bedL2_0	0.42	0.80 (0.06)	1.90 (0.14)	7.02×10^{-10}
H3K27ac_Hnisz.bedL2_0	0.39	0.76 (0.06)	1.94 (0.15)	2.13×10^{-9}
Repressed_Hoffman.extend.500.bedL2_0	0.72	0.44 (0.05)	0.61 (0.07)	1.10×10^{-8}
SuperEnhancer_Hnisz.extend.500.bedL2_0	0.17	0.36 (0.03)	2.12 (0.19)	1.76×10^{-8}
H3K4me1_Trynka.extend.500.bedL2_0	0.61	0.97 (0.06)	1.60 (0.10)	3.30×10^{-8}
Conserved_LindbladToh.extend.500.bedL2_0	0.33	0.73 (0.07)	2.21 (0.22)	6.08×10^{-8}
H3K9ac_Trynka.bedL2_0	0.13	0.59 (0.09)	4.72 (0.69)	1.31×10^{-7}
Conserved_LindbladToh.bedL2_0	0.03	0.40 (0.07)	15.58 (2.85)	2.90×10^{-7}
H3K4me3_Trynka.extend.500.bedL2_0	0.26	0.65 (0.07)	2.53 (0.29)	6.61×10^{-7}
SuperEnhancer_Hnisz.bedL2_0	0.17	0.34 (0.04)	2.05 (0.21)	9.01×10^{-7}
H3K4me1_Trynka.bedL2_0	0.42	0.98 (0.11)	2.31 (0.26)	1.36×10^{-6}
H3K27ac_PGC2.extend.500.bedL2_0	0.34	0.66 (0.07)	1.98 (0.22)	1.14×10^{-5}
Enhancer_Hoffman.extend.500.bedL2_0	0.09	0.35 (0.06)	3.91 (0.67)	2.30×10^{-5}
DHS_peaks_Trynka.bedL2_0	0.11	0.67 (0.13)	6.05 (1.20)	5.30×10^{-5}
DGF_ENCODE.extend.500.bedL2_0	0.54	0.89 (0.09)	1.65 (0.16)	1.54×10^{-4}
H3K4me1_peaks_Trynka.bedL2_0	0.17	0.71 (0.14)	4.17 (0.85)	2.50×10^{-4}
TFBS_ENCODE.bedL2_0	0.13	0.59 (0.12)	4.46 (0.94)	3.96×10^{-4}
FetalDHS_Trynka.extend.500.bedL2_0	0.28	0.62 (0.09)	2.17 (0.32)	4.17×10^{-4}
Enhancer_Hoffman.bedL2_0	0.04	0.30 (0.07)	7.16 (1.75)	4.23×10^{-4}
FetalDHS_Trynka.bedL2_0	0.08	0.52 (0.13)	6.25 (1.51)	5.62×10^{-4}
H3K27ac_PGC2.bedL2_0	0.27	0.57 (0.09)	2.11 (0.34)	1.24×10^{-3}
DHS_Trynka.bedL2_0	0.17	0.64 (0.16)	3.85 (0.94)	2.75×10^{-3}
Intron_UCSC.extend.500.bedL2_0	0.4	0.49 (0.04)	1.23 (0.09)	0.010
Coding_UCSC.bedL2_0	0.01	0.12 (0.04)	8.40 (3.07)	0.016
TFBS_ENCODE.extend.500.bedL2_0	0.34	0.60 (0.11)	1.74 (0.31)	0.017

H3K4me3_peaks_Trynka.bedL2_0	0.04	0.25 (0.09)	5.91 (2.08)	0.018
H3K9ac_peaks_Trynka.bedL2_0	0.04	0.25 (0.09)	6.57 (2.35)	0.019
UTR_3_UCSC.bedL2_0	0.01	0.08 (0.03)	6.73 (2.71)	0.034
H3K4me3_Trynka.bedL2_0	0.13	0.32 (0.09)	2.37 (0.68)	0.040
PromoterFlanking_Hoffman.bedL2_0	0.01	-0.06 (0.03)	-7.00 (4.13)	0.054
TSS_Hoffman.extend.500.bedL2_0	0.03	0.11 (0.04)	3.25 (1.19)	0.063
Repressed_Hoffman.bedL2_0	0.46	0.21 (0.13)	0.47 (0.29)	0.066
Coding_UCSC.extend.500.bedL2_0	0.06	0.15 (0.05)	2.28 (0.71)	0.068
DHS_Trynka.extend.500.bedL2_0	0.5	0.69 (0.11)	1.39 (0.22)	0.079
UTR_5_UCSC.extend.500.bedL2_0	0.03	0.08 (0.03)	3.13 (1.24)	0.082
DGF_ENCODE.bedL2_0	0.14	0.36 (0.14)	2.63 (1.02)	0.108
TSS_Hoffman.bedL2_0	0.02	0.09 (0.05)	4.95 (2.54)	0.126
Intron_UCSC.bedL2_0	0.39	0.45 (0.04)	1.17 (0.11)	0.139
Promoter_UCSC.extend.500.bedL2_0	0.04	0.09 (0.03)	2.30 (0.90)	0.142
Enhancer_Andersson.extend.500.bedL2_0	0.02	0.07 (0.03)	3.49 (1.83)	0.179
Transcribed_Hoffman.extend.500.bedL2_0	0.76	0.69 (0.06)	0.91 (0.08)	0.242
CTCF_Hoffman.bedL2_0	0.02	0.08 (0.06)	3.56 (2.55)	0.318
CTCF_Hoffman.extend.500.bedL2_0	0.07	0.01 (0.06)	0.14 (0.91)	0.345
WeakEnhancer_Hoffman.bedL2_0	0.02	0.07 (0.06)	3.34 (2.81)	0.409
Enhancer_Andersson.bedL2_0	0	0.03 (0.03)	6.06 (7.09)	0.475
WeakEnhancer_Hoffman.extend.500.bedL2_0	0.09	0.13 (0.06)	1.47 (0.73)	0.522
Promoter_UCSC.bedL2_0	0.03	0.05 (0.04)	1.58 (1.33)	0.663
Transcribed_Hoffman.bedL2_0	0.35	0.32 (0.12)	0.92 (0.34)	0.824
UTR_5_UCSC.bedL2_0	0.01	0.01 (0.02)	1.37 (4.55)	0.935
PromoterFlanking_Hoffman.extend.500.bedL2_0	0.03	0.04 (0.04)	1.09 (1.28)	0.946
UTR_3_UCSC.extend.500.bedL2_0	0.03	0.03 (0.03)	1.03 (1.17)	0.976
baseL2_0	1	1.00 (0.00)	1.00 (0.00)	-

SE, standard error; Prop_SNPs, proportion of SNPs in category; Prop_h2, proportion of heritability attributed to category

Table 4. Summary of datasets for analysis in this study.

Study/Project	Dataset	Description	Analysis
ENCODE Project			
Artery	Open Chromatin		
	ENCSR000EIH	aortic smooth muscle cell	Aorta SKAT analyses groupings
	ENCSR000EMC	aortic adventitia cell	Aorta SKAT analyses groupings
	ENCSR000EOG	pulmonary artery endothelial cell	Aorta SKAT analyses groupings
	ENCSR000EOH	pulmonary artery fibroblast	Aorta SKAT analyses groupings
	ENCSR630REB	tibial artery	Aorta, Tibial Artery SKAT analyses groupings
	Histone Modification		
	ENCSR519CFV	aorta H3K27Ac	Aorta SKAT analyses groupings
	ENCSR015GFK	aorta H3K27Ac	Aorta SKAT analyses groupings
	ENCSR318HUC	aorta H3K27Ac	Aorta SKAT analyses groupings
	ENCSR069UMW	aorta H3K27Ac	Aorta SKAT analyses groupings
	ENCSR322TJD	aorta H3K27Ac	Aorta SKAT analyses groupings
	ENCSR233LCT	tibial artery H3K4me1	Tibial Artery SKAT analyses groupings
Kidney	Open Chromatin		
	ENCSR000EOK	renal cortical epithelial cell	Kidney SKAT analyses groupings
	ENCSR000EOM	glomerular endothelial cell	Kidney SKAT analyses groupings
	ENCSR000EPW	proximal tubule epithelial cell	Kidney SKAT analyses groupings
	ENCSR785BDQ	glomerular visceral epithelial cell (3yo)	Kidney SKAT analyses groupings
GERA	Genotypes, BP phenotype	71404 European-Ancestry individuals	SKAT, BP traits
	Summary statistics	80,792 European-Ancestry individuals	Partitioned heritability analyses, MetaXcan
ARIC	Genotypes, QT interval	9,083 European-Ancestry individuals	SKAT, QT interval

	Summary statistics	9,083 European-Ancestry individuals	Partitioned heritability analyses, MetaXcan
GTEx	Genotypes, Expression		
	Aorta	197 samples	SKAT, expression (for BP)
	Tibial Artery	285 samples	SKAT, expression (for BP)
	Heart Atrial Appendage	159 samples	SKAT, expression (for QT)
	Heart Left Ventricle	190 samples	SKAT, expression (for QT)

Table 5. SCN5A SKAT and MetaXcan results for QT interval

Tissue	N.qt	p.qt. dsvm	p.qt.eq	p.qt.def	N. GTEx	p.GTEx. dsvm	p.GTEx. eq	p.GTEx. def	N.MetX	p.MetX
Heart	39	4.86×10^{-4}	1.65×10^{-4}	0.015	-	-	-	-	-	-
Heart.LV	-	-	-	-	34	0.857	0.405	0.461	-	-
Heart.AA	-	-	-	-	34	0.265	0.546	0.310	13	0.011
Aorta_Artery	-	-	-	-	-	-	-	-	-	-
Tibial_Artery	2	0.682	0.383	0.675	-	-	-	-	-	-
ENCSR000EOK	3	0.051	0.218	0.154	-	-	-	-	-	-
ENCSR000EOM	6	0.077	1.28×10^{-3}	0.099	-	-	-	-	-	-
ENCSR000EPW	5	0.076	0.100	0.099	-	-	-	-	-	-
ENCSR785BDQ	3	0.215	0.335	0.204	-	-	-	-	-	-

N.qt, number of variants analyzed in SKAT analysis of QT in ARIC
p.qt.*, p-values from SKAT analysis of QT in ARIC with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTEx.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

LV, left ventricle; AA, atrial appendage; ENCSR000EOK, renal cortical epithelial cell; ENCSR000EOM, glomerular endothelial cell; ENCSR000EPW, epithelial cell of proximal tubule; ENCSR785BDQ, glomerular visceral epithelial cell

Table 6. NOS1AP SKAT and MetaXcan results for QT interval

Tissue	N.qt	p.qt.dsvm	p.qt.eq	p.qt.def	N.GTEx	p.GTEx.dsvm	p.GTEx.eq	p.GTEx.def	N. MetX	p. MetX
Heart	69	3.53×10^{-15}	3.89×10^{-18}	1.90×10^{-4}	-	-	-	-	-	-
Heart.LV	-	-	-	-	59	3.85×10^{-4}	1.51×10^{-4}	0.054	-	-
Heart.AA	-	-	-	-	59	0.181	0.182	0.315	-	-
Aorta_Artery	18	1.09×10^{-6}	1.34×10^{-17}	0.026	12	0.789	0.949	0.999	-	-
Tibial_Artery	20	1.94×10^{-18}	6.70×10^{-22}	2.48×10^{-4}	12	2.23×10^{-3}	0.012	0.264	-	-
ENCSR000EOK	43	4.87×10^{-10}	5.66×10^{-19}	1.47×10^{-5}	-	-	-	-	-	-
ENCSR000EOM	28	7.79×10^{-15}	4.39×10^{-20}	3.29×10^{-5}	-	-	-	-	-	-
ENCSR000EPW	33	1.69×10^{-22}	1.63×10^{-21}	2.73×10^{-6}	-	-	-	-	-	-
ENCSR785BDQ	16	2.81×10^{-21}	3.17×10^{-14}	0.012	-	-	-	-	-	-

N.qt, number of variants analyzed in SKAT analysis of QT in ARIC
p.qt.*, p-values from SKAT analysis of QT in ARIC with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTEx.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

LV, left ventricle; AA, atrial appendage; ENCSR000EOK, renal cortical epithelial cell; ENCSR000EOM, glomerular endothelial cell; ENCSR000EPW, epithelial cell of proximal tubule

Table 7. Aorta SBP SKAT and MetaXcan results

Gene	N. sbp	p.sbp. dsvm	p.sbp. eq	p.sbp. def	N. GTEx	p.GTEx. dsvm	p.GTEx. eq	p.GTEx. def	N.MetX	p.MetX	Analysis
NPR3	33	3.73×10^{-7}	1.31×10^{-6}	0.570	17	0.911	0.871	0.929	-	-	SKAT
WBP1L	39	3.14×10^{-6}	5.91×10^{-7}	0.085	37	3.30×10^{-4}	4.76×10^{-4}	0.137	21	1.35×10^{-5}	SKAT
ARL3	22	7.31×10^{-4}	4.94×10^{-5}	3.61×10^{-3}	14	0.177	0.070	0.414	4	2.96×10^{-9}	MetaXcan
FURIN	5	0.234	0.074	0.176	14	9.16×10^{-5}	1.02×10^{-4}	0.780	3	1.27×10^{-7}	MetaXcan
ARHGAP42	12	0.032	$6.25 \times 10\text{-}^{3}$	0.424	13	8.36×10^{-3}	7.40×10^{-3}	0.112	35	2.08×10^{-7}	MetaXcan
TMEM133	2	0.422	0.193	0.454	2	0.167	0.202	0.170	49	2.39×10^{-7}	MetaXcan
IDH2	7	0.026	0.021	0.779	7	0.070	1.19×10^{-3}	0.685	10	1.46×10^{-6}	MetaXcan
ATF1	4	0.064	5.06×10^{-3}	0.077	3	1.05×10^{-4}	1.03×10^{-9}	3.00×10^{-3}	31	2.48×10^{-6}	MetaXcan
NSF	-	-	-	-	-	-	-	-	27	4.08×10^{-6}	MetaXcan

N.sbp, number of variants analyzed in SKAT analysis of SBP in GERA
p.sbp.*, p-values from SKAT analysis of BP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTex.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

Table 8. Aorta DBP SKAT and MetaXcan results

Gene	N. dbp	p.dbp.dsvm	p.dbp.eq	p.dbp.def	N. GTEx	p.GTEx.dsvm	p.GTEx.eq	p.GTEx.def	N. MetX	p. MetX	Analysis
ATXN2	9	1.99×10^{-13}	9.77×10^{-11}	0.148	11	0.988	0.987	0.937	-	-	SKAT
NOV	10	2.40×10^{-8}	1.98×10^{-8}	0.664	7	4.54×10^{-5}	4.04×10^{-5}	1	99	0.102	SKAT
SH2B3	8	4.49×10^{-8}	3.37×10^{-10}	0.521	9	0.415	0.368	0.988	-	-	SKAT
COX14	6	7.58×10^{-8}	7.49×10^{-7}	0.138	4	1.01×10^{-3}	5.97×10^{-5}	0.338	5	2.77×10^{-3}	SKAT
CERS5	7	7.66×10^{-8}	1.92×10^{-6}	0.138	5	0.151	0.012	0.494	-	-	SKAT
RP4-605O3.4	7	7.66×10^{-8}	1.92×10^{-6}	0.138	5	0.018	5.11×10^{-4}	0.425	-	-	SKAT
SMARCD1	5	3.12×10^{-7}	3.10×10^{-6}	0.138	4	0.353	0.243	0.308	-	-	SKAT
GPD1	5	3.12×10^{-7}	3.10×10^{-6}	0.138	4	0.101	0.251	0.772	-	-	SKAT
IGFBP3	4	3.27×10^{-7}	7.67×10^{-8}	6.47×10^{-3}	3	0.011	0.014	0.842	-	-	SKAT
CEP170	6	6.41×10^{-7}	1.49×10^{-6}	8.44×10^{-7}	7	1.91×10^{-4}	1.72×10^{-4}	3.59×10^{-4}	19	1.69×10^{-3}	SKAT
CSK	3	9.49×10^{-7}	8.91×10^{-7}	0.068	3	2.67×10^{-5}	2.49×10^{-5}	0.579	17	7.93×10^{-5}	SKAT
ULK3	2	9.93×10^{-7}	9.76×10^{-7}	0.826	3	8.35×10^{-4}	7.02 x 10-4	0.164	9	2.09×10^{-4}	SKAT
SLX4IP	70	1.14×10^{-6}	7.91×10^{-8}	0.198	69	0.497	0.294	0.167	-	-	SKAT

TMEM133	2	0.308	0.369	0.305	2	0.167	0.202	0.170	49	3.74×10^{-7}	MetaXcan
ATF1	4	0.048	2.56×10^{-3}	0.020	3	1.05×10^{-4}	1.03×10^{-9}	3.00×10^{-3}	31	8.59×10^{-7}	MetaXcan
HIST1H2AE	51	0.015	9.33×10^{-3}	0.429	22	1.82×10^{-3}	1.48×10^{-3}	0.085	13	1.13×10^{-6}	MetaXcan
ARHGAP42	12	0.083	6.59×10^{-3}	0.389	13	8.36×10^{-3}	7.40×10^{-3}	0.112	35	1.29×10^{-6}	MetaXcan
SPATA2L	10	9.51×10^{-3}	0.028	0.684	11	6.72×10^{-4}	1.03×10^{-3}	0.928	10	7.70×10^{-6}	MetaXcan
ULK4	5	3.27×10^{-8}	5.04×10^{-8}	1.48×10^{-3}	5	2.67×10^{-22}	2.95×10^{-22}	0.094	42	2.95×10^{-10}	SKAT/ MetaXcan
SDCCAG8	10	6.71×10^{-7}	1.74×10^{-6}	0.359	10	1.99×10^{-6}	2.26×10^{-8}	0.481	9	2.95×10^{-9}	SKAT/ MetaXcan

N.dbp, number of variants analyzed in SKAT analysis of DBP in GERA
p.dbp.*, p-values from SKAT analysis of DBP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTex.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

Table 9. Tibial Artery SBP SKAT and MetaXcan results

Gene	N. sbp	p.sbp. dsvm	p.sbp.eq	p.sbp. def	N. GTEx	p.GTEx .dsvm	p.GTEx .eq	p.GTEx. def	N.MetX	p. MetX	Analysis
MTHFR	24	5.06×10^{-9}	6.54×10^{-8}	2.62×10^{-4}	28	7.31×10^{-8}	1.13×10^{-5}	0.161	35	0.076	SKAT
AGTRAP	15	1.26×10^{-8}	4.58×10^{-9}	1.54×10^{-4}	19	0.462	0.584	0.922	1	0.098	SKAT
C10orf32	8	6.73×10^{-8}	2.08×10^{-8}	4.36×10^{-3}	7	1.87×10^{-14}	2.05×10^{-18}	7.30×10^{-8}	-	-	SKAT
CYP17A1	15	6.78×10^{-8}	3.60×10^{-8}	0.031	14	0.343	0.394	1	-	-	SKAT
CYP17A1-AS1	15	6.78×10^{-8}	3.60×10^{-8}	0.031	14	0.029	0.056	0.796	-	-	SKAT
NPR3	21	8.47×10^{-7}	1.05×10^{-6}	0.713	13	0.041	0.091	0.128	-	-	SKAT
KCNK3	15	1.75×10^{-6}	3.12×10^{-6}	9.82×10^{-3}	13	0.942	0.918	0.693	-	-	SKAT
CNNM2	30	1.75×10^{-6}	3.01×10^{-4}	3.51×10^{-5}	32	0.261	0.507	0.860	-	-	SKAT
PPM1L	13	3.35×10^{-6}	3.18×10^{-6}	0.742	11	0.873	0.839	0.633	-	-	SKAT
ATP2B1	4	0.452	0.511	0.348	4	1	0.882	0.915	6	5.82×10^{-12}	MetaXcan
TMEM133	-	-	-	-	-	-	-	-	37	2.87×10^{-9}	MetaXcan
ARHGAP42	19	2.18×10^{-3}	3.39×10^{-3}	0.579	20	4.57×10^{-7}	3.14×10^{-7}	0.208	38	1.73×10^{-8}	MetaXcan
ARL3	19	2.61×10^{-3}	1.33×10^{-4}	3.91×10^{-5}	15	0.432	0.117	0.063	5	3.54×10^{-7}	MetaXcan

AP1B1	5	6.61×10^{-3}	0.096	0.552	3	0.684	0.297	0.760	1	3.80×10^{-7}	MetaXcan
THBS2	32	0.031	0.034	0.403	11	6.21×10^{-4}	1.03×10^{-3}	0.834	36	9.70×10^{-7}	MetaXcan
ARL17A	-	-	-	-	-	-	-	-	55	1.28×10^{-6}	MetaXcan
MST1	2	0.048	8.10×10^{-3}	0.051	1	0.015	0.015	0.015	8	4.12×10^{-6}	MetaXcan
NUDT13	4	5.60×10^{-3}	8.30×10^{-4}	9.80×10^{-3}	4	9.61×10^{-10}	3.59×10^{-22}	3.22×10^{-9}	46	4.94×10^{-6}	MetaXcan
NME6	6	2.52×10^{-4}	2.52×10^{-4}	0.961	2	4.60×10^{-3}	6.33×10^{-15}	0.868	26	6.27×10^{-6}	MetaXcan
SCAMP5	7	6.27×10^{-4}	3.21×10^{-4}	0.574	5	1.25×10^{-11}	1.36×10^{-6}	5.09×10^{-5}	35	6.56×10^{-6}	MetaXcan
CLCN6	24	5.06×10^{-9}	6.54×10^{-8}	2.62×10^{-4}	28	3.58×10^{-7}	6.14×10^{-8}	3.58×10^{-6}	10	2.67×10^{-9}	SKAT/ MetaXcan

N.sbp, number of variants analyzed in SKAT analysis of SBP in GERA
p.sbp.*, p-values from SKAT analysis of BP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTex.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

Table 10. Tibial Artery DBP SKAT and MetaXcan results

Gene	N. dbp	p.dbp. dsvm	p.dbp.eq	p.dbp. def	N. GTEx	p.GTEx. dsvm	p.GTEx .eq	p.GTEx. def	N. MetX	p. MetX	Analysis
ATXN2	7	6.63×10^{-11}	4.46×10^{-11}	0.227	7	0.221	0.196	0.619	-	-	SKAT
SH2B3	9	1.22×10^{-10}	8.46×10^{-11}	0.485	9	0.129	0.170	0.370	-	-	SKAT
IGFBP3	10	3.35×10^{-7}	2.98×10^{-7}	0.299	6	0.154	0.387	0.638	-	-	SKAT
CERS5	6	1.65×10^{-6}	3.75×10^{-6}	3.54×10^{-3}	2	3.61×10^{-4}	3.90×10^{-4}	3.55 x 10-4	33	1.17×10^{-4}	SKAT
COX14	6	1.65×10^{-6}	3.75×10^{-6}	3.54×10^{-3}	2	2.56×10^{-5}	1.62×10^{-5}	6.54 x 10-4	33	3.67×10^{-3}	SKAT
RP4-605O3.4	6	1.65×10^{-6}	3.75×10^{-6}	3.54×10^{-3}	2	1.63×10^{-7}	1.76×10^{-7}	2.36 x 10-7	-	-	SKAT
HECTD4	2	0.015	0.016	0.077	1	0.860	0.860	0.860	4	1.46×10^{-9}	MetaXcan
TMEM133	-	-	-	-	-	-	-	-	37	4.35×10^{-8}	MetaXcan
SPATA2L	10	0.015	0.015	0.095	12	1.03×10^{-4}	3.79×10^{-5}	0.815	2	6.38×10^{-8}	MetaXcan
SCAMP5	7	5.46×10^{-5}	7.75×10^{-5}	0.331	5	1.25×10^{-11}	1.36×10^{-6}	5.09 x 10-5	35	8.04×10^{-8}	MetaXcan
SDCCAG8	6	6.33×10^{-5}	5.15×10^{-5}	0.130	7	4.29×10^{-5}	1.73×10^{-5}	0.273	17	8.08×10^{-8}	MetaXcan
ARHGAP42	19	0.021	0.028	0.092	20	4.57×10^{-7}	3.14×10^{-7}	0.208	38	5.06×10^{-7}	MetaXcan
AP1B1	5	0.039	0.096	0.634	3	0.684	0.297	0.760	1	3.40×10^{-6}	MetaXcan

CDK11A	-	-	-	-	-	-	-	-	9	4.82×10^{-6}	MetaXcan
ATP2B1	4	0.879	0.481	0.894	4	1	0.882	0.915	6	5.18×10^{-6}	MetaXcan
NOV	17	1.29×10^{-8}	1.15×10^{-8}	0.731	16	6.91×10^{-6}	7.95×10^{-6}	0.943	34	1.44×10^{-6}	SKAT/ MetaXcan
ULK4	12	1.06×10^{-6}	2.60×10^{-10}	8.12×10^{-3}	11	3.34×10^{-13}	3.61×10^{-19}	0.158	50	5.40×10^{-12}	SKAT/ MetaXcan

N.dbp, number of variants analyzed in SKAT analysis of DBP in GERA
p.dbp.*, p-values from SKAT analysis of DBP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights
N.GTEx, number of variants analyzed in SKAT analysis of expression in GTEx
p.GTex.*, p-values from SKAT analysis of expression in GTEx with deltasvm (dsvm), eq (equal), or default (def) weights
N.MetX, number of variants used in MetaXcan prediction results
p.MetX, MetaXcan p-value

Table 11. Aorta individual variants analyzed from six genes

Gene	SNP	MAF	deltaSVM	SBP BETA (SE)	P.sbp	DBP BETA (SE)	P.dbp	GTEx.slope (SE)	GTEx.p. nominal
CLCN6	rs55867221	0.098	0.854	-0.589 (0.11)	4.59×10^{-8}	-0.235 (0.07)	7.64×10^{-4}	0.205 (0.08)	0.013
CLCN6	rs2151655	0.292	-0.301	0.021 (0.07)	0.769	-0.034 (0.05)	0.458	-0.019 (0.05)	0.723
CLCN6	rs2151654	0.098	-0.167	-0.591 (0.11)	4.14×10^{-8}	-0.235 (0.07)	7.60×10^{-4}	0.223 (0.08)	6.85×10^{-3}
CLCN6	rs59375726	0.173	-0.258	0.059 (0.09)	0.492	-0.013 (0.06)	0.808	0.052 (0.07)	0.456
CLCN6	rs3737966	0.387	-0.451	-0.241 (0.07)	3.18×10^{-4}	-0.142 (0.04)	1.04×10^{-3}	-0.118 (0.05)	0.022
CLCN6	rs3820192	0.098	-0.712	-0.595 (0.11)	3.47×10^{-8}	-0.237 (0.07)	6.78×10^{-4}	0.235 (0.08)	6.38×10^{-3}
CLCN6	rs1537516	0.097	-0.552	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.251 (0.08)	3.57×10^{-3}
CLCN6	rs1537515	0.097	0.252	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.251 (0.08)	3.57×10^{-3}
CLCN6	rs1537514	0.098	-0.674	-0.593 (0.11)	3.92×10^{-8}	-0.237 (0.07)	6.96×10^{-4}	0.223 (0.08)	6.83×10^{-3}
CLCN6	rs55686944	0.048	-0.209	0.066 (0.15)	0.660	0.107 (0.10)	0.272	-0.030 (0.11)	0.788
CLCN6	rs1572151	0.071	-0.346	0.225 (0.12)	0.072	0.129 (0.08)	0.111	-0.042 (0.09)	0.652
CLCN6	rs17421511	0.168	1.143	0.082 (0.09)	0.337	0.011 (0.06)	0.848	0.031 (0.07)	0.663
CLCN6	rs45449597	0.166	0.062	0.070 (0.09)	0.416	0.009 (0.06)	0.866	0.031 (0.07)	0.664
CLCN6	rs45608437	0.167	-1.696	0.068 (0.09)	0.429	0.008 (0.06)	0.881	0.033 (0.07)	0.640
CLCN6	rs4846052	0.414	-0.368	-0.192 (0.07)	3.42×10^{-3}	-0.126 (0.04)	2.86×10^{-3}	-0.124 (0.05)	0.015
CLCN6	rs17037388	0.158	-0.060	-0.496	1.46 x	-0.260	4.40×10^{-6}	0.261 (0.07)	2.29×10^{-4}

				(0.09)	10^{-8}	(0.06)			
CLCN6	rs3753588	0.098	-0.852	-0.607 (0.11)	1.33 x 10^{-8}	-0.238 (0.07)	5.72 x 10^{-4}	0.240 (0.09)	7.58 x 10^{-3}
CLCN6	rs17367629	0.098	-0.085	-0.607 (0.11)	1.32 x 10^{-8}	-0.239 (0.07)	5.54 x 10^{-4}	0.240 (0.09)	7.58 x 10^{-3}
CLCN6	rs13306560	0.056	-1.018	-0.208 (0.14)	0.134	-0.243 (0.09)	6.74 x 10^{-3}	0.242 (0.12)	0.047
CLCN6	rs45460391	0.002	-0.210	0.026 (0.68)	0.970	0.049 (0.44)	0.912	-	-
CLCN6	rs111334345	0.039	-0.624	-0.418 (0.16)	0.011	-0.380 (0.11)	3.38 x 10^{-4}	0.275 (0.14)	0.044
CLCN6	rs6699618	0.154	0.822	-0.510 (0.09)	6.16 x 10^{-9}	-0.266 (0.06)	2.83 x 10^{-6}	0.250 (0.07)	7.14 x 10^{-4}
CLCN6	rs75747410	0.045	-0.609	-0.290 (0.15)	0.057	-0.136 (0.10)	0.169	0.659 (0.14)	4.90 x 10^{-6}
CLCN6	rs198411	0.140	-0.364	0.048 (0.09)	0.603	-0.027 (0.06)	0.646	0.059 (0.07)	0.419
CLCN6	rs198412	0.083	1.260	0.239 (0.12)	0.037	0.126 (0.07)	0.092	0.005 (0.08)	0.949
CLCN6	rs198413	0.140	-0.479	0.050 (0.09)	0.584	-0.029 (0.06)	0.621	0.059 (0.07)	0.419
CLCN6	rs198414	0.140	0.415	0.054 (0.09)	0.561	-0.013 (0.06)	0.829	0.059 (0.07)	0.419
CLCN6	rs17350396	0.169	-0.675	0.072 (0.09)	0.398	0.010 (0.06)	0.863	0.021 (0.07)	0.774
SDCCAG8	rs6662991	0.369	-0.335	0.151 (0.07)	0.025	0.176 (0.04)	5.41 x 10^{-5}	0.311 (0.07)	1.09 x 10^{-5}
SDCCAG8	rs3904682	0.483	0.138	0.130 (0.06)	0.043	0.169 (0.04)	4.98 x 10^{-5}	-0.396 (0.06)	7.53 x 10^{-9}
SDCCAG8	rs28599724	0.468	0.820	0.095 (0.07)	0.146	0.164 (0.04)	1.12 x 10^{-4}	-0.378 (0.07)	6.79 x 10^{-8}
SDCCAG8	rs10926978	0.323	1.826	0.195 (0.07)	4.37 x 10^{-3}	0.219 (0.04)	7.17 x 10^{-7}	0.311 (0.07)	1.31 x 10^{-5}
SDCCAG8	rs6682448	0.323	-1.065	0.194 (0.07)	4.56 x 10^{-3}	0.217 (0.04)	9.07 x 10^{-7}	0.307 (0.07)	1.49 x 10^{-5}
SDCCAG8	rs10926981	0.463	-0.518	0.129 (0.06)	0.043	0.188 (0.04)	5.89 x 10^{-6}	-0.395 (0.06)	5.21 x 10^{-9}

SDCCAG8	rs112223723	0.033	-0.373	-0.066 (0.18)	0.711	0.139 (0.12)	0.229	-0.128 (0.18)	0.482
SDCCAG8	rs3006933	0.424	0.239	-0.092 (0.07)	0.158	0.001 (0.04)	0.984	0.110 (0.08)	0.149
SDCCAG8	rs113544916	0.010	0.342	0.466 (0.32)	0.146	0.158 (0.21)	0.448	-	-
SDCCAG8	rs1058305	0.258	0.669	-0.062 (0.07)	0.393	0.049 (0.05)	0.301	0.197 (0.08)	0.019
NOV	rs2447155	0.182	-0.133	0.117 (0.08)	0.158	-0.229 (0.05)	1.78×10^{-5}	-0.222 (0.05)	4.08×10^{-5}
NOV	rs75254116	0.018	-0.353	0.094 (0.24)	0.700	0.065 (0.16)	0.680	-	-
NOV	rs115372142	0.219	-0.587	0.054 (0.08)	0.482	-0.281 (0.05)	1.68×10^{-8}	0.230 (0.05)	3.32×10^{-5}
NOV	rs16892527	0.018	0.016	0.070 (0.24)	0.772	0.036 (0.16)	0.817	-	-
NOV	rs16892528	0.018	-0.582	0.070 (0.24)	0.772	0.036 (0.16)	0.817	-	-
NOV	rs17791184	0.220	0.372	0.050 (0.08)	0.516	-0.279 (0.05)	1.94×10^{-8}	0.230 (0.05)	3.51×10^{-5}
NOV	rs73711267	0.238	0.613	0.057 (0.07)	0.447	-0.261 (0.05)	6.45×10^{-8}	0.229 (0.05)	2.84×10^{-5}
NOV	rs61064779	0.220	0.108	0.050 (0.08)	0.512	-0.281 (0.05)	1.65×10^{-8}	0.230 (0.05)	3.32×10^{-5}
NOV	rs144538053	0.006	-0.956	0.640 (0.43)	0.138	0.323 (0.28)	0.247	-	-
NOV	rs17188373	0.220	2.089	0.047 (0.08)	0.537	-0.281 (0.05)	1.52×10^{-8}	0.230 (0.05)	3.32×10^{-5}
C10orf32	rs284862	0.432	0.283	-0.254 (0.06)	9.11×10^{-5}	0.000 (0.04)	1.000	0.372 (0.06)	6.07×10^{-10}
C10orf32	rs284861	0.121	0.145	-0.672 (0.10)	7.62×10^{-12}	-0.201 (0.06)	1.57 x 10-3	0.036 (0.09)	0.706
C10orf32	rs17115073	0.019	0.759	0.184 (0.23)	0.427	0.194 (0.15)	0.196	0.097 (0.17)	0.573
C10orf32	rs284858	0.411	-0.798	-0.285 (0.07)	1.25×10^{-5}	-0.020 (0.04)	0.643	0.392 (0.06)	1.15×10^{-10}

C10orf32	rs34102073	0.015	0.255	0.197 (0.26)	0.445	0.192 (0.17)	0.252	-	-
C10orf32	rs284857	0.420	-0.768	-0.294 (0.07)	6.23×10^{-6}	-0.035 (0.04)	0.401	0.403 (0.06)	4.28×10^{-11}
C10orf32	rs284856	0.411	0.421	-0.290 (0.07)	8.52×10^{-6}	-0.023 (0.04)	0.588	0.403 (0.06)	4.28×10^{-11}
C10orf32	rs284855	0.411	-0.830	-0.289 (0.07)	8.99×10^{-6}	-0.023 (0.04)	0.592	0.403 (0.06)	4.28×10^{-11}
C10orf32	rs284854	0.432	-0.034	-0.271 (0.06)	2.81×10^{-5}	-0.007 (0.04)	0.875	0.371 (0.06)	6.40×10^{-10}
C10orf32	rs284853	0.122	-0.230	-0.670 (0.10)	7.32×10^{-12}	-0.208 (0.06)	1.03×10^{-3}	0.036 (0.09)	0.699
C10orf32	rs14849	0.019	-0.297	0.217 (0.23)	0.347	0.215 (0.15)	0.148	0.103 (0.17)	0.550
C10orf32	rs6892	0.176	0.621	0.201 (0.08)	0.017	0.072 (0.05)	0.185	0.228 (0.09)	0.012
C10orf32	rs17523050	0.012	-0.821	0.626 (0.29)	0.032	0.205 (0.19)	0.277	-	-
C10orf32	rs3824754	0.097	0.253	-0.726 (0.11)	1.40×10^{-11}	-0.193 (0.07)	5.39×10^{-3}	-0.092 (0.12)	0.457
C10orf32	rs12416687	0.253	-0.731	0.035 (0.07)	0.631	0.112 (0.05)	0.019	-0.695 (0.05)	3.80×10^{-26}
C10orf32	rs17881215	0.111	-0.637	-0.041 (0.10)	0.687	0.001 (0.07)	0.985	-0.576 (0.09)	1.87×10^{-9}
C10orf32	rs3740400	0.370	0.247	-0.238 (0.07)	3.14×10^{-4}	0.027 (0.04)	0.535	-0.485 (0.05)	6.22×10^{-17}
ULK4	rs11564435	0.007	0.223	-0.891 (0.39)	0.021	-0.462 (0.25)	0.064	-	-
ULK4	rs11564463	0.023	0.310	-0.472 (0.21)	0.027	-0.266 (0.14)	0.054	-	-
ULK4	rs1795330	0.140	0.064	-0.056 (0.09)	0.539	-0.012 (0.06)	0.845	-0.046 (0.10)	0.651
ULK4	rs7624877	0.026	0.226	0.167 (0.20)	0.406	0.368 (0.13)	4.73×10^{-3}	0.311 (0.17)	0.064
ULK4	rs73830585	0.150	0.944	-0.162 (0.09)	0.068	0.334 (0.06)	6.64×10^{-9}	0.923 (0.06)	1.72×10^{-32}

MTHFR	rs6668392	0.009	-1.023	-0.037 (0.34)	0.912	0.205 (0.22)	0.347	-	-
MTHFR	rs55867221	0.098	0.854	-0.589 (0.11)	4.59×10^{-8}	-0.235 (0.07)	7.64×10^{-4}	0.219 (0.10)	0.028
MTHFR	rs2151655	0.292	-0.301	0.021 (0.07)	0.769	-0.034 (0.05)	0.458	-0.163 (0.06)	0.011
MTHFR	rs2151654	0.098	-0.167	-0.591 (0.11)	4.14×10^{-8}	-0.235 (0.07)	7.60×10^{-4}	0.233 (0.10)	0.020
MTHFR	rs59375726	0.173	-0.258	0.059 (0.09)	0.492	-0.013 (0.06)	0.808	0.336 (0.08)	4.59×10^{-5}
MTHFR	rs3737966	0.387	-0.451	-0.241 (0.07)	3.18×10^{-4}	-0.142 (0.04)	1.04×10^{-3}	-0.226 (0.06)	2.45×10^{-4}
MTHFR	rs3820192	0.098	-0.712	-0.595 (0.11)	3.47×10^{-8}	-0.237 (0.07)	6.78×10^{-4}	0.210 (0.10)	0.044
MTHFR	rs1537516	0.097	-0.552	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.213 (0.10)	0.042
MTHFR	rs1537515	0.097	0.252	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.213 (0.10)	0.042
MTHFR	rs1537514	0.098	-0.674	-0.593 (0.11)	3.92×10^{-8}	-0.237 (0.07)	6.96×10^{-4}	0.233 (0.10)	0.020
MTHFR	rs55686944	0.048	-0.209	0.066 (0.15)	0.660	0.107 (0.10)	0.272	0.021 (0.13)	0.872
MTHFR	rs1572151	0.071	-0.346	0.225 (0.12)	0.072	0.129 (0.08)	0.111	-0.026 (0.11)	0.816
MTHFR	rs17421511	0.168	1.143	0.082 (0.09)	0.337	0.011 (0.06)	0.848	0.334 (0.08)	7.24×10^{-5}
MTHFR	rs45449597	0.166	0.062	0.070 (0.09)	0.416	0.009 (0.06)	0.866	0.334 (0.08)	7.60×10^{-5}
MTHFR	rs45608437	0.167	-1.696	0.068 (0.09)	0.429	0.008 (0.06)	0.881	0.335 (0.08)	7.09×10^{-5}
MTHFR	rs4846052	0.414	-0.368	-0.192 (0.07)	3.42×10^{-3}	-0.126 (0.04)	2.86×10^{-3}	-0.238 (0.06)	8.22×10^{-5}
MTHFR	rs17037388	0.158	-0.060	-0.496 (0.09)	1.46×10^{-8}	-0.260 (0.06)	4.40×10^{-6}	0.236 (0.09)	6.37×10^{-3}
MTHFR	rs3753588	0.098	-0.852	-0.607	1.33 x	-0.238	5.72×10^{-4}	0.164 (0.11)	0.133

				(0.11)	10^{-8}	(0.07)			
MTHFR	rs17367629	0.098	-0.085	-0.607 (0.11)	1.32 x 10^{-8}	-0.239 (0.07)	5.54 x 10^{-4}	0.164 (0.11)	0.133
MTHFR	rs13306560	0.056	-1.018	-0.208 (0.14)	0.134	-0.243 (0.09)	6.74 x 10^{-3}	0.303 (0.15)	0.039
MTHFR	rs45460391	0.002	-0.210	0.026 (0.68)	0.970	0.049 (0.44)	0.912	-	-
MTHFR	rs111334345	0.039	-0.624	-0.418 (0.16)	0.011	-0.380 (0.11)	3.38 x 10^{-4}	0.373 (0.16)	0.023
MTHFR	rs6699618	0.154	0.822	-0.510 (0.09)	6.16 x 10^{-9}	-0.266 (0.06)	2.83 x 10^{-6}	0.244 (0.09)	6.64 x 10^{-3}
MTHFR	rs75747410	0.045	-0.609	-0.290 (0.15)	0.057	-0.136 (0.10)	0.169	0.183 (0.18)	0.308
MTHFR	rs198411	0.140	-0.364	0.048 (0.09)	0.603	-0.027 (0.06)	0.646	0.009 (0.09)	0.922
MTHFR	rs198412	0.083	1.260	0.239 (0.12)	0.037	0.126 (0.07)	0.092	-0.057 (0.10)	0.566
MTHFR	rs198413	0.140	-0.479	0.050 (0.09)	0.584	-0.029 (0.06)	0.621	0.009 (0.09)	0.922
MTHFR	rs198414	0.140	0.415	0.054 (0.09)	0.561	-0.013 (0.06)	0.829	0.009 (0.09)	0.922
MTHFR	rs17350396	0.169	-0.675	0.072 (0.09)	0.398	0.010 (0.06)	0.863	0.316 (0.08)	2.16 x 10^{-4}

SE, standard error; GTEx.slope, GTEx beta from the GTEx authors' FastQTL analyses; GTEx.p.nominal, nominal p-value from GTEx authors' FastQTL analyses

Table 12. Tibial artery individual variants analyzed from six genes

Gene	SNP	MAF	deltaSVM	SBP BETA (SE)	P.sbp	DBP BETA (SE)	P.dbp	GTEx.slope (SE)	GTEx.p. nominal
CLCN6	rs55867221	0.098	0.919	-0.589 (0.11)	4.59×10^{-8}	-0.235 (0.07)	7.64×10^{-4}	0.297 (0.06)	2.58×10^{-7}
CLCN6	rs183631107	0.010	-0.316	-0.663 (0.33)	0.045	-0.189 (0.21)	0.376	-	-
CLCN6	rs2184226	0.074	-0.135	0.175 (0.12)	0.151	0.113 (0.08)	0.153	-0.044 (0.07)	0.521
CLCN6	rs3737967	0.042	-0.081	-0.306 (0.16)	0.055	-0.115 (0.10)	0.266	0.375 (0.09)	4.88×10^{-5}
CLCN6	rs1537516	0.097	0.144	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.319 (0.06)	1.49×10^{-7}
CLCN6	rs1537515	0.097	-0.389	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.319 (0.06)	1.49×10^{-7}
CLCN6	rs1537514	0.098	-0.541	-0.593 (0.11)	3.92×10^{-8}	-0.237 (0.07)	6.96×10^{-4}	0.306 (0.06)	1.26×10^{-7}
CLCN6	rs2274976	0.042	-0.262	-0.318 (0.16)	0.046	-0.123 (0.10)	0.233	0.375 (0.09)	4.88×10^{-5}
CLCN6	rs55686944	0.048	0.919	0.066 (0.15)	0.660	0.107 (0.10)	0.272	-0.053 (0.08)	0.524
CLCN6	rs6541005	0.438	0.581	-0.202 (0.07)	2.80×10^{-3}	-0.125 (0.04)	4.35×10^{-3}	-0.139 (0.04)	4.12×10^{-4}
CLCN6	rs6541006	0.068	0.000	-0.194 (0.13)	0.138	-0.205 (0.08)	0.016	-0.015 (0.07)	0.833
CLCN6	rs45577533	0.055	-0.084	-0.236 (0.14)	0.092	-0.241 (0.09)	7.73×10^{-3}	0.179 (0.09)	0.039
CLCN6	rs17037388	0.158	0.233	-0.496	$1.46 \times$	-0.260	$4.40 \times$	0.235	1.42×10^{-6}

				(0.09)	10^{-8}	(0.06)	10^{-6}	(0.05)	
CLCN6	rs2066470	0.098	-0.321	-0.605 (0.11)	1.63 x 10^{-8}	-0.239 (0.07)	5.54 x 10^{-4}	0.326 (0.06)	9.46 x 10^{-8}
CLCN6	rs3753588	0.098	-0.325	-0.607 (0.11)	1.33 x 10^{-8}	-0.238 (0.07)	5.72 x 10^{-4}	0.329 (0.06)	5.78 x 10^{-8}
CLCN6	rs13306561	0.155	-0.413	-0.504 (0.09)	9.34 x 10^{-9}	-0.259 (0.06)	5.38 x 10^{-6}	0.267 (0.05)	4.14 x 10^{-8}
CLCN6	rs45594036	0.056	-0.565	-0.235 (0.14)	0.091	-0.251 (0.09)	5.37 x 10^{-3}	0.162 (0.09)	0.085
CLCN6	rs111334345	0.039	-0.883	-0.418 (0.16)	0.011	-0.380 (0.11)	3.38 x 10^{-4}	0.180 (0.11)	0.108
CLCN6	rs3737965	0.045	-0.310	-0.289 (0.15)	0.058	-0.137 (0.10)	0.165	0.364 (0.09)	7.75 x 10^{-5}
CLCN6	rs198411	0.140	-0.029	0.048 (0.09)	0.603	-0.027 (0.06)	0.646	-0.057 (0.05)	0.260
CLCN6	rs198412	0.083	0.045	0.239 (0.12)	0.037	0.126 (0.07)	0.092	-0.132 (0.06)	0.029
CLCN6	rs198413	0.140	-0.528	0.050 (0.09)	0.584	-0.029 (0.06)	0.621	-0.057 (0.05)	0.260
CLCN6	rs198414	0.140	0.399	0.054 (0.09)	0.561	-0.013 (0.06)	0.829	-0.057 (0.05)	0.260
CLCN6	rs17350396	0.169	-0.813	0.072 (0.09)	0.398	0.010 (0.06)	0.863	0.063 (0.05)	0.232
SDCCAG8	rs7539093	0.016	-0.089	0.389 (0.26)	0.128	0.272 (0.17)	0.100	0.156 (0.15)	0.298
SDCCAG8	rs12124014	0.496	0.945	-0.150 (0.06)	0.019	-0.172 (0.04)	3.17 x 10^{-5}	-0.276 (0.06)	1.78 x 10^{-5}
SDCCAG8	rs3904682	0.483	0.120	0.130 (0.06)	0.043	0.169 (0.04)	4.98 x 10^{-5}	-0.283 (0.06)	6.40 x 10^{-6}
SDCCAG8	rs61833906	0.122	0.870	-0.151	0.122	-0.043	0.496	-0.064	0.467

				(0.10)		(0.06)		(0.09)	
SDCCAG8	rs12125812	0.106	-0.227	-0.023 (0.11)	0.829	-0.042 (0.07)	0.551	-0.261 (0.10)	7.36×10^{-3}
SDCCAG8	rs146560296	0.003	-0.431	-0.644 (0.60)	0.281	-0.227 (0.39)	0.558	-	-
NOV	rs56013326	0.218	-0.153	0.048 (0.08)	0.538	-0.281 (0.05)	1.76×10^{-8}	0.229 (0.05)	3.84×10^{-6}
NOV	rs2279112	0.266	-0.736	0.091 (0.07)	0.207	-0.234 (0.05)	4.71×10^{-7}	0.187 (0.04)	3.79×10^{-5}
NOV	rs16892524	0.018	-0.558	0.088 (0.24)	0.716	0.047 (0.16)	0.763	0.117 (0.17)	0.499
NOV	rs16892525	0.018	-0.955	0.071 (0.24)	0.769	0.039 (0.16)	0.804	0.117 (0.17)	0.496
NOV	rs57788046	0.220	-0.579	0.051 (0.08)	0.510	-0.280 (0.05)	1.68×10^{-8}	0.227 (0.05)	4.52×10^{-6}
NOV	rs17791208	0.222	-0.168	0.039 (0.08)	0.610	-0.282 (0.05)	1.68×10^{-8}	0.221 (0.05)	5.94×10^{-6}
NOV	rs17187482	0.220	0.220	0.050 (0.08)	0.512	-0.281 (0.05)	1.65×10^{-8}	0.227 (0.05)	4.52×10^{-6}
NOV	rs16892556	0.030	-0.186	0.052 (0.19)	0.783	-0.042 (0.12)	0.731	0.004 (0.12)	0.976
NOV	rs72684301	0.030	-0.143	0.052 (0.19)	0.783	-0.042 (0.12)	0.731	0.000 (0.12)	0.998
NOV	rs58210708	0.249	0.390	0.051 (0.07)	0.491	-0.264 (0.05)	2.77×10^{-8}	0.208 (0.05)	1.19×10^{-5}
NOV	rs117920888	0.014	0.076	0.375 (0.27)	0.166	0.188 (0.18)	0.285	-	-
NOV	rs56782665	0.220	-0.501	0.048 (0.08)	0.534	-0.281 (0.05)	1.50×10^{-8}	0.227 (0.05)	4.52×10^{-6}
NOV	rs144538053	0.006	0.216	0.640	0.138	0.323	0.247	-	-

				(0.43)		(0.28)			
NOV	rs17188373	0.220	2.183	0.047 (0.08)	0.537	-0.281 (0.05)	1.52 x 10^{-8}	0.227 (0.05)	4.52 x 10^{-6}
NOV	rs3928531	0.220	-0.178	0.047 (0.08)	0.537	-0.281 (0.05)	1.52 x 10^{-8}	0.227 (0.05)	4.52 x 10^{-6}
NOV	rs1461696	0.253	0.710	0.048 (0.07)	0.511	-0.263 (0.05)	2.43 x 10^{-8}	0.202 (0.05)	2.30 x 10^{-5}
NOV	rs1461697	0.033	-0.281	0.033 (0.18)	0.852	-0.052 (0.11)	0.648	-0.047 (0.12)	0.693
C10orf32	rs284856	0.411	0.840	-0.290 (0.07)	8.52 x 10^{-6}	-0.023 (0.04)	0.588	0.384 (0.05)	4.09 x 10^{-14}
C10orf32	rs284855	0.411	-0.670	-0.289 (0.07)	8.99 x 10^{-6}	-0.023 (0.04)	0.592	0.383 (0.05)	3.60 x 10^{-14}
C10orf32	rs284854	0.432	0.029	-0.271 (0.06)	2.81 x 10^{-5}	-0.007 (0.04)	0.875	0.372 (0.05)	1.05 x 10^{-13}
C10orf32	rs284853	0.122	-0.737	-0.670 (0.10)	7.32 x 10^{-12}	-0.208 (0.06)	1.03 x 10^{-3}	-0.104 (0.08)	0.178
C10orf32	rs142463040	0.012	0.653	0.184 (0.30)	0.535	-0.001 (0.19)	0.994	0.185 (0.19)	0.331
C10orf32	rs17523050	0.012	-0.756	0.626 (0.29)	0.032	0.205 (0.19)	0.277	-	-
C10orf32	rs3824754	0.097	0.381	-0.726 (0.11)	1.40 x 10^{-11}	-0.193 (0.07)	5.39 x 10^{-3}	0.053 (0.10)	0.588
C10orf32	rs17881215	0.111	0.254	-0.041 (0.10)	0.687	0.001 (0.07)	0.985	-0.575 (0.07)	6.61 x 10^{-15}
ULK4	rs78048770	0.008	1.414	-0.191 (0.35)	0.589	-0.245 (0.23)	0.283	-	-
ULK4	rs11564437	0.047	-2.505	-0.201 (0.15)	0.186	-0.115 (0.10)	0.241	0.000 (0.14)	0.999
ULK4	rs1795313	0.489	0.589	0.014	0.831	-0.012	0.774	-0.004	0.947

				(0.06)		(0.04)		(0.06)	
ULK4	rs9851971	0.340	-0.721	0.055 (0.07)	0.418	-0.023 (0.04)	0.601	-0.047 (0.05)	0.372
ULK4	rs35391137	0.086	0.795	-0.139 (0.11)	0.223	-0.122 (0.07)	0.099	0.054 (0.10)	0.577
ULK4	rs28702390	0.167	-0.478	-0.107 (0.09)	0.208	0.358 (0.06)	7.98×10^{-11}	0.631 (0.06)	3.88×10^{-20}
ULK4	rs6791806	0.172	0.285	-0.112 (0.08)	0.185	0.361 (0.05)	3.75×10^{-11}	0.645 (0.06)	1.13×10^{-23}
ULK4	rs73081321	0.150	1.087	-0.151 (0.09)	0.089	0.332 (0.06)	8.88×10^{-9}	0.655 (0.06)	1.54×10^{-20}
ULK4	rs73071261	0.150	-0.183	-0.162 (0.09)	0.068	0.334 (0.06)	6.64×10^{-9}	0.643 (0.06)	1.00×10^{-19}
ULK4	rs9868854	0.018	-0.432	0.234 (0.24)	0.331	0.426 (0.16)	6.33×10^{-3}	0.155 (0.20)	0.443
ULK4	rs74745987	0.018	-0.579	0.111 (0.24)	0.645	0.383 (0.16)	0.014	-	-
ULK4	rs78703690	0.018	-0.157	0.117 (0.24)	0.626	0.385 (0.15)	0.013	-	-
MTHFR	rs55867221	0.098	0.919	-0.589 (0.11)	4.59×10^{-8}	-0.235 (0.07)	7.64×10^{-4}	0.189 (0.07)	4.58×10^{-3}
MTHFR	rs183631107	0.010	-0.316	-0.663 (0.33)	0.045	-0.189 (0.21)	0.376	-	-
MTHFR	rs2184226	0.074	-0.135	0.175 (0.12)	0.151	0.113 (0.08)	0.153	0.038 (0.08)	0.624
MTHFR	rs3737967	0.042	-0.081	-0.306 (0.16)	0.055	-0.115 (0.10)	0.266	0.145 (0.11)	0.172
MTHFR	rs1537516	0.097	0.144	-0.599 (0.11)	2.96×10^{-8}	-0.236 (0.07)	7.43×10^{-4}	0.185 (0.07)	8.54×10^{-3}
MTHFR	rs1537515	0.097	-0.389	-0.599	2.96 x	-0.236	7.43 x	0.185	8.54×10^{-3}

				(0.11)	10^{-8}	(0.07)	10^{-4}	(0.07)	
MTHFR	rs1537514	0.098	-0.541	-0.593 (0.11)	3.92 x 10^{-8}	-0.237 (0.07)	6.96 x 10^{-4}	0.197 (0.07)	3.35 x 10^{-3}
MTHFR	rs2274976	0.042	-0.262	-0.318 (0.16)	0.046	-0.123 (0.10)	0.233	0.145 (0.11)	0.172
MTHFR	rs55686944	0.048	0.919	0.066 (0.15)	0.660	0.107 (0.10)	0.272	-0.014 (0.10)	0.885
MTHFR	rs6541005	0.438	0.581	-0.202 (0.07)	2.80 x 10^{-3}	-0.125 (0.04)	4.35 x 10^{-3}	-0.182 (0.04)	4.54 x 10^{-5}
MTHFR	rs6541006	0.068	0.000	-0.194 (0.13)	0.138	-0.205 (0.08)	0.016	0.085 (0.08)	0.281
MTHFR	rs45577533	0.055	-0.084	-0.236 (0.14)	0.092	-0.241 (0.09)	7.73 x 10^{-3}	0.129 (0.10)	0.190
MTHFR	rs17037388	0.158	0.233	-0.496 (0.09)	1.46 x 10^{-8}	-0.260 (0.06)	4.40 x 10^{-6}	0.069 (0.06)	0.222
MTHFR	rs2066470	0.098	-0.321	-0.605 (0.11)	1.63 x 10^{-8}	-0.239 (0.07)	5.54 x 10^{-4}	0.156 (0.07)	0.028
MTHFR	rs3753588	0.098	-0.325	-0.607 (0.11)	1.33 x 10^{-8}	-0.238 (0.07)	5.72 x 10^{-4}	0.137 (0.07)	0.052
MTHFR	rs13306561	0.155	-0.413	-0.504 (0.09)	9.34 x 10^{-9}	-0.259 (0.06)	5.38 x 10^{-6}	0.065 (0.06)	0.252
MTHFR	rs45594036	0.056	-0.565	-0.235 (0.14)	0.091	-0.251 (0.09)	5.37 x 10^{-3}	0.066 (0.11)	0.540
MTHFR	rs111334345	0.039	-0.883	-0.418 (0.16)	0.011	-0.380 (0.11)	3.38 x 10^{-4}	0.201 (0.13)	0.113
MTHFR	rs3737965	0.045	-0.310	-0.289 (0.15)	0.058	-0.137 (0.10)	0.165	0.044 (0.11)	0.677
MTHFR	rs198411	0.140	-0.029	0.048 (0.09)	0.603	-0.027 (0.06)	0.646	-0.114 (0.06)	0.044
MTHFR	rs198412	0.083	0.045	0.239	0.037	0.126	0.092	-0.082	0.236

				(0.12)		(0.07)		(0.07)	
MTHFR	rs198413	0.140	-0.528	0.050 (0.09)	0.584	-0.029 (0.06)	0.621	-0.114 (0.06)	0.044
MTHFR	rs198414	0.140	0.399	0.054 (0.09)	0.561	-0.013 (0.06)	0.829	-0.114 (0.06)	0.044
MTHFR	rs17350396	0.169	-0.813	0.072 (0.09)	0.398	0.010 (0.06)	0.863	0.290 (0.06)	7.38×10^{-7}

SE, standard error; GTEx.slope, GTEx beta from the GTEx authors' FastQTL analyses; GTEx.p.nominal, nominal p-value from GTEx authors' FastQTL analyses

Table 13. Kidney SBP SKAT and MetaXcan results

Gene	Experiment	N.sbp	p.sbp.dsvm	p.sbp.eq	p.sbp.def
BSND	ENCSR000EOK	13	0.791	0.796	0.611
CASR	ENCSR000EOK	21	0.586	0.794	0.605
CLCNKA	ENCSR000EOK	8	0.491	0.502	0.426
CLCNKB	ENCSR000EOK	9	0.521	0.382	0.454
CUL3	ENCSR000EOK	18	0.251	0.388	0.049
CYP11B1	ENCSR000EOK	12	0.019	0.016	1.20×10^{-3}
CYP11B2	ENCSR000EOK	10	0.013	0.014	2.91×10^{-3}
CYP17A1	ENCSR000EOK	7	0.064	1.11×10^{-6}	0.018
HSD11B2	ENCSR000EOK	4	0.121	0.104	0.248
KCNJ1	ENCSR000EOK	13	0.993	0.958	0.789
KCNJ5	ENCSR000EOK	11	0.548	0.742	0.982
KLHL3	ENCSR000EOK	12	0.573	0.431	0.345
NR3C2	ENCSR000EOK	26	0.317	0.270	0.148
SCNN1A	ENCSR000EOK	7	0.472	0.452	0.767
SCNN1B	ENCSR000EOK	7	0.049	0.310	0.479
SCNN1G	ENCSR000EOK	8	0.187	0.276	1
SLC12A1	ENCSR000EOK	6	1	0.909	0.713
SLC12A3	ENCSR000EOK	16	0.316	0.168	0.491
WNK1	ENCSR000EOK	11	0.366	0.557	0.637
WNK4	ENCSR000EOK	1	0.010	0.010	0.010
BSND	ENCSR000EOM	12	0.887	0.737	0.562
CASR	ENCSR000EOM	2	0.868	0.790	0.672
CLCNKA	ENCSR000EOM	10	0.678	0.657	0.540
CLCNKB	ENCSR000EOM	10	0.678	0.657	0.540
CUL3	ENCSR000EOM	8	0.060	0.250	0.044
CYP11B1	ENCSR000EOM	4	0.019	0.018	0.281
CYP11B2	ENCSR000EOM	3	0.015	0.015	0.281
CYP17A1	ENCSR000EOM	11	1.26×10^{-7}	2.88×10^{-7}	0.013
HSD11B2	ENCSR000EOM	6	0.029	0.044	0.595
KCNJ1	ENCSR000EOM	9	0.091	0.554	0.060
KCNJ5	ENCSR000EOM	6	0.083	0.150	0.163
KLHL3	ENCSR000EOM	12	0.067	0.012	0.208
NR3C2	ENCSR000EOM	10	0.254	0.126	0.149
SCNN1A	ENCSR000EOM	5	0.518	0.572	0.727
SCNN1B	ENCSR000EOM	5	0.822	0.930	0.873
SCNN1G	ENCSR000EOM	5	0.374	0.298	0.540
SLC12A1	ENCSR000EOM	4	0.703	0.896	0.704
SLC12A3	ENCSR000EOM	12	0.056	0.077	0.245
WNK1	ENCSR000EOM	7	0.153	0.158	0.864
WNK4	ENCSR000EOM	1	0.010	0.010	0.010
BSND	ENCSR000EPW	10	0.665	0.702	0.591
CASR	ENCSR000EPW	11	0.638	0.798	0.335

CLCNKA	ENCSR000EPW	3	0.276	0.203	0.433
CLCNKB	ENCSR000EPW	4	0.231	0.176	0.464
CUL3	ENCSR000EPW	8	0.033	0.195	0.039
CYP11B1	ENCSR000EPW	3	0.015	0.015	0.281
CYP11B2	ENCSR000EPW	5	0.015	0.017	0.488
CYP17A1	ENCSR000EPW	9	0.012	9.92×10^{-7}	0.045
HSD11B2	ENCSR000EPW	1	0.051	0.051	0.051
KCNJ1	ENCSR000EPW	7	0.951	0.954	0.411
KCNJ5	ENCSR000EPW	10	0.522	0.699	0.417
KLHL3	ENCSR000EPW	8	0.395	0.525	0.428
NR3C2	ENCSR000EPW	24	0.252	0.189	0.027
SCNN1A	ENCSR000EPW	6	0.496	0.360	0.767
SCNN1B	ENCSR000EPW	7	0.141	0.181	0.482
SCNN1G	ENCSR000EPW	6	0.725	0.565	1
SLC12A1	ENCSR000EPW	4	1	0.896	0.704
SLC12A3	ENCSR000EPW	13	0.163	0.116	0.358
WNK1	ENCSR000EPW	11	0.555	0.515	0.054
WNK4	ENCSR000EPW	1	0.010	0.010	0.010
BSND	ENCSR785BDQ	10	0.908	0.854	0.668
CASR	ENCSR785BDQ	1	0.465	0.465	0.465
CLCNKA	ENCSR785BDQ	10	0.470	0.606	0.682
CLCNKB	ENCSR785BDQ	11	0.465	0.476	0.695
CUL3	ENCSR785BDQ	2	0.737	0.314	0.982
CYP11B1	ENCSR785BDQ	4	0.383	0.353	0.026
CYP11B2	ENCSR785BDQ	2	0.022	0.026	0.996
CYP17A1	ENCSR785BDQ	12	7.24×10^{-3}	3.93×10^{-5}	0.020
HSD11B2	ENCSR785BDQ	6	0.164	0.067	0.061
KCNJ1	ENCSR785BDQ	7	0.985	0.879	0.400
KCNJ5	ENCSR785BDQ	2	1	1	1
KLHL3	ENCSR785BDQ	10	0.594	0.126	0.888
NR3C2	ENCSR785BDQ	9	0.299	0.290	0.037
SCNN1A	ENCSR785BDQ	9	0.525	0.700	0.980
SCNN1B	ENCSR785BDQ	4	0.887	1	0.983
SCNN1G	ENCSR785BDQ	6	0.727	0.565	1
SLC12A1	ENCSR785BDQ	2	0.844	0.848	0.848
SLC12A3	ENCSR785BDQ	28	0.212	0.440	0.666
WNK1	ENCSR785BDQ	7	0.259	0.414	0.128
WNK4	ENCSR785BDQ	-	-	-	-

N.sbp, number of variants analyzed in SKAT analysis of SBP in GERA
p.sbp.*, p-values from SKAT analysis of DBP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights

ENCSR000EOK, renal cortical epithelial cell; ENCSR000EOM, glomerular endothelial cell; ENCSR000EPW, epithelial cell of proximal tubule; ENCSR785BDQ, glomerular visceral epithelial cell

Table 14. Kidney DBP SKAT and MetaXcan results

Gene	Experiment	N.dbp	p.dbp.dsvm	p.dbp.eq	p.dbp.def
BSND	ENCSR000EOK	13	0.933	0.831	0.061
CASR	ENCSR000EOK	21	0.987	0.986	0.856
CLCNKA	ENCSR000EOK	8	0.155	0.283	0.108
CLCNKB	ENCSR000EOK	9	0.159	0.410	0.137
CUL3	ENCSR000EOK	18	0.284	0.183	0.855
CYP11B1	ENCSR000EOK	12	0.102	0.095	0.013
CYP11B2	ENCSR000EOK	10	0.091	0.094	0.023
CYP17A1	ENCSR000EOK	7	0.505	0.189	0.337
HSD11B2	ENCSR000EOK	4	0.030	0.023	0.114
KCNJ1	ENCSR000EOK	13	0.966	0.879	0.398
KCNJ5	ENCSR000EOK	11	0.716	0.598	0.487
KLHL3	ENCSR000EOK	12	0.823	0.922	0.622
NR3C2	ENCSR000EOK	26	0.731	0.766	0.104
SCNN1A	ENCSR000EOK	7	0.715	0.621	0.599
SCNN1B	ENCSR000EOK	7	3.25×10^{-4}	0.012	0.053
SCNN1G	ENCSR000EOK	8	0.095	0.027	0.461
SLC12A1	ENCSR000EOK	6	0.334	0.259	0.860
SLC12A3	ENCSR000EOK	16	0.106	0.072	0.175
WNK1	ENCSR000EOK	11	0.365	0.300	0.442
WNK4	ENCSR000EOK	1	0.828	0.828	0.828
BSND	ENCSR000EOM	12	0.638	0.776	0.052
CASR	ENCSR000EOM	2	0.884	0.825	0.759
CLCNKA	ENCSR000EOM	10	0.294	0.546	0.133
CLCNKB	ENCSR000EOM	10	0.294	0.546	0.133
CUL3	ENCSR000EOM	8	0.270	0.266	0.798
CYP11B1	ENCSR000EOM	4	0.130	0.129	0.689
CYP11B2	ENCSR000EOM	3	0.101	0.105	0.689
CYP17A1	ENCSR000EOM	11	0.044	0.076	0.187
HSD11B2	ENCSR000EOM	6	0.015	0.012	0.387
KCNJ1	ENCSR000EOM	9	0.098	0.317	0.101
KCNJ5	ENCSR000EOM	6	0.105	0.499	0.598
KLHL3	ENCSR000EOM	12	0.900	0.924	0.515
NR3C2	ENCSR000EOM	10	0.612	0.302	0.072
SCNN1A	ENCSR000EOM	5	0.783	0.709	0.547
SCNN1B	ENCSR000EOM	5	0.792	0.907	0.913
SCNN1G	ENCSR000EOM	5	0.196	0.085	0.920
SLC12A1	ENCSR000EOM	4	0.107	0.244	0.865
SLC12A3	ENCSR000EOM	12	7.17×10^{-3}	0.044	0.234
WNK1	ENCSR000EOM	7	0.492	0.589	1
WNK4	ENCSR000EOM	1	0.828	0.828	0.828
BSND	ENCSR000EPW	10	0.849	0.769	0.062
CASR	ENCSR000EPW	11	0.921	0.845	0.872

CLCNKA	ENCSR000EPW	3	0.435	0.438	0.367
CLCNKB	ENCSR000EPW	4	0.718	0.639	0.413
CUL3	ENCSR000EPW	8	0.225	0.085	0.783
CYP11B1	ENCSR000EPW	3	0.101	0.105	0.689
CYP11B2	ENCSR000EPW	5	0.096	0.102	0.599
CYP17A1	ENCSR000EPW	9	0.370	0.185	0.262
HSD11B2	ENCSR000EPW	1	5.56×10^{-3}	5.56×10^{-3}	5.56×10^{-3}
KCNJ1	ENCSR000EPW	7	0.796	0.873	0.366
KCNJ5	ENCSR000EPW	10	0.673	0.546	0.190
KLHL3	ENCSR000EPW	8	0.819	0.867	0.448
NR3C2	ENCSR000EPW	24	0.495	0.374	0.032
SCNN1A	ENCSR000EPW	6	0.738	0.446	0.599
SCNN1B	ENCSR000EPW	7	0.080	0.017	0.053
SCNN1G	ENCSR000EPW	6	0.631	0.061	0.297
SLC12A1	ENCSR000EPW	4	0.403	0.244	0.865
SLC12A3	ENCSR000EPW	13	0.024	0.050	0.217
WNK1	ENCSR000EPW	11	0.297	0.173	0.031
WNK4	ENCSR000EPW	1	0.828	0.828	0.828
BSND	ENCSR785BDQ	10	0.673	0.492	0.062
CASR	ENCSR785BDQ	1	0.967	0.967	0.967
CLCNKA	ENCSR785BDQ	10	0.296	0.457	0.268
CLCNKB	ENCSR785BDQ	11	0.299	0.568	0.303
CUL3	ENCSR785BDQ	2	0.184	0.040	0.248
CYP11B1	ENCSR785BDQ	4	0.152	0.134	2.67×10^{-3}
CYP11B2	ENCSR785BDQ	2	0.095	0.107	0.508
CYP17A1	ENCSR785BDQ	12	0.587	0.218	0.355
HSD11B2	ENCSR785BDQ	6	0.046	7.76×10^{-3}	3.39×10^{-3}
KCNJ1	ENCSR785BDQ	7	0.957	0.795	0.128
KCNJ5	ENCSR785BDQ	2	1	1	0.875
KLHL3	ENCSR785BDQ	10	0.785	0.778	0.424
NR3C2	ENCSR785BDQ	9	0.789	0.784	0.019
SCNN1A	ENCSR785BDQ	9	0.780	0.949	0.925
SCNN1B	ENCSR785BDQ	4	0.310	0.445	0.555
SCNN1G	ENCSR785BDQ	6	0.176	0.061	0.297
SLC12A1	ENCSR785BDQ	2	0.279	0.282	0.282
SLC12A3	ENCSR785BDQ	28	0.037	0.415	0.393
WNK1	ENCSR785BDQ	7	0.799	0.416	0.079
WNK4	ENCSR785BDQ	-	-	-	-

N.dbp, number of variants analyzed in SKAT analysis of DBP in GERA
p.dbp.*, p-values from SKAT analysis of DBP in GERA with deltasvm (dsvm), eq (equal), or default (def) weights

ENCSR000EOK, renal cortical epithelial cell; ENCSR000EOM, glomerular endothelial cell; ENCSR000EPW, epithelial cell of proximal tubule; ENCSR785BDQ, glomerular visceral epithelial cell

Table 15. Individual variants analyzed in kidney for *C10orf32* and *CYP17A1*

SNP	MAF	SBP BETA (SE)	P.sbp	DBP BETA (SE)	P.dbp	deltaSVM. ENCSR000EOK	deltaSVM. ENCSR000EOM	deltaSVM. ENCSR000EPW	deltaSVM. ENCSR785BDQ	Gene
rs1046778	0.305	-0.223 (0.07)	1.24×10^{-3}	0.045 (0.04)	0.319	1.690	1.185	1.518	-	C10orf32
rs117325917	0.016	0.022 (0.26)	0.931	0.046 (0.17)	0.782	-	-	-	0.117	CYP17A1
rs2253703	0.387	-0.124 (0.07)	0.057	-0.027 (0.04)	0.515	-	-	-	0.864	CYP17A1
rs2253709	0.288	0.179 (0.07)	0.012	0.017 (0.05)	0.719	-	-	-	-1.653	CYP17A1
rs141778169	0.026	0.277 (0.20)	0.172	0.178 (0.13)	0.175	-0.225	0.462	-0.138	-0.403	C10orf32/ CYP17A1
rs142463040	0.012	0.184 (0.30)	0.535	-0.001 (0.19)	0.994	0.174	0.358	0.291	-	C10orf32/ CYP17A1
rs17115125	0.004	0.454 (0.49)	0.358	-0.171 (0.32)	0.592	-0.471	-	-0.485	-0.811	C10orf32/ CYP17A1
rs17523050	0.012	0.626 (0.29)	0.032	0.205 (0.19)	0.277	-0.415	-0.414	-0.616	0.266	C10orf32/ CYP17A1
rs2486758	0.210	0.134 (0.08)	0.088	0.032 (0.05)	0.532	-1.625	-	-0.794	0.317	C10orf32/ CYP17A1
rs3824754	0.097	-0.726 (0.11)	1.40×10^{-11}	-0.193 (0.07)	5.39×10^{-3}	0.211	0.434	0.279	0.038	C10orf32/ CYP17A1
rs743572	0.394	-0.231 (0.07)	3.98E-04	0.014 (0.04)	0.735	-0.132	-	0.092	0.429	C10orf32/ CYP17A1
rs14849	0.019	0.217 (0.23)	0.347	0.215 (0.15)	0.148	-	-0.031	-	-	C10orf32/ CYP17A1
rs284853	0.122	-0.670	7.32 x	-0.208	1.03 x	-	-0.311	-	-	C10orf32/

		(0.10)	10^{-12}	(0.06)	10^{-3}					CYP17A1
rs284854	0.432	-0.271 (0.06)	2.81 x 10^{-5}	-0.007 (0.04)	0.875	-	0.421	-	-	C10orf32/ CYP17A1
rs284855	0.411	-0.289 (0.07)	8.99 x 10^{-6}	-0.023 (0.04)	0.592	-	-0.268	-	-	C10orf32/ CYP17A1
rs284856	0.411	-0.290 (0.07)	8.52 x 10^{-6}	-0.023 (0.04)	0.588	-	0.146	-	-	C10orf32/ CYP17A1
rs6892	0.176	0.201 (0.08)	0.017	0.072 (0.05)	0.185	-	0.205	-	-	C10orf32/ CYP17A1
rs743575	0.296	0.038 (0.07)	0.585	0.098 (0.05)	0.031	-	0.385	-	-	C10orf32/ CYP17A1
rs182622705	0.026	0.277 (0.20)	0.172	0.178 (0.13)	0.175	-	-	0.058	-	C10orf32/ CYP17A1
rs190942123	0.033	0.252 (0.18)	0.164	0.131 (0.12)	0.262	-	-	-1.394	-	C10orf32/ CYP17A1
rs12416687	0.253	0.035 (0.07)	0.631	0.112 (0.05)	0.019	-	-	-	-0.568	C10orf32/ CYP17A1
rs17881215	0.111	-0.041 (0.10)	0.687	0.001 (0.07)	0.985	-	-	-	-0.830	C10orf32/ CYP17A1
rs3740400	0.370	-0.238 (0.07)	3.14 x 10^{-4}	0.027 (0.04)	0.535	-	-	-	-0.152	C10orf32/ CYP17A1

SE, standard error

ENCSR000EOK, renal cortical epithelial cell; ENCSR000EOM, glomerular endothelial cell; ENCSR000EPW, epithelial cell of proximal tubule; ENCSR785BDQ, glomerular visceral epithelial cell

Chapter 5: Conclusions

Hypertension is a complex multigenic disease affected by many genetic as well as environmental factors. While the renal physiology of the genes involved in monogenic forms of hypertension and hypotension has been studied considerably, their contributions to the overall incidence of hypertension is unknown. Additionally, most of the common variants with small effects at hundreds of loci associated with blood pressure (BP) remain to be characterized at the level of genes and variants in relevant tissues. Therefore, the first part of this dissertation has focused on the exome, identifying rare variants in genes of interest (Chapters 2 and 3); the second part (Chapter 4) focused on tissues of interest and putative regulatory regions, identifying variants of all frequencies as well as specific genes, using tissue-specific information.

The focus of the study in Chapter 2 was to explain more of the phenotypic variation in BP traits and to advance blood pressure genetic findings in a population historically underrepresented in blood pressure genetic studies. We genotyped 15,914 African-ancestry individuals for functionally relevant rare and low-frequency variants and identified rare variants in 10 genes. The major limitation in this study was that the array was designed primarily from sequences of European-ancestry individuals, which most likely contribute to the rare allele frequency of the significant results. Nevertheless, a literature search confirmed cardiovascular and related roles for several of these genes and they are of interest, especially in this population, moving forward with larger sample sizes.

Chapter 3 detailed a replication analysis aimed at identifying deleterious rare variants in three hypotension genes, *SLC12A1*, *SLC12A3*, and *KCNJ1*, in order to determine if a BP-lowering effect could be observed in a new study. If successful, that would give way to conducting similar analyses to identify genes in which rare variants are responsible for blood pressure variation in the general population. The original study was a targeted sequencing study in the Framingham Heart Study (FHS) offspring cohort,[10] while the replication effort was carried out in the ARIC study with exome sequencing data. We observed that, while we could not replicate the effect using all variants meeting the original study criteria for loss-of-function variants, we did observe similar effects when studying only the subset of variants from the original study also appearing in the ARIC data. Several studies following up the FHS study provided biochemical evidence of loss-of-function for the FHS variants, while we did not have functional follow-up of any of our variants. These results suggest that variants at specific sites could be driving the effect; however, due to the presence of sequencing or other errors, we may have not been able to identify the cumulative effect of true causal variants due to dilution from the presence of false variants. Variants with extremely low allele counts are often discarded in sequencing analyses, but this study indicates that such effects can still be observed and be biologically relevant, and that with improved filtering, and eventually highly accurate sequencing, it will become possible to identify all such variants. Furthermore, the evidence suggests the utility of this approach in identifying possibly novel genes with rare loss-of-function variants associated with large blood pressure effects in the general population.

Renal salt handling mechanisms regulating BP have been well characterized, but less is known about other pathways involved in BP regulation. In Chapter 4, we identified the aorta and tibial arteries as two tissues of interest for BP GWAS loci, and subsequently identified several genes and variants at GWAS loci with strong expression support in these tissues. We also examined the 20 monogenic BP syndrome genes in several kidney cell types, their suggested tissue of action, for which open chromatin data were available, in order to determine the contributions of regulatory variants for these genes. These results were largely inconclusive, in no small part due to the limited data available on gene regulation in the kidney. However, these analyses revealed a cell-type-specificity of *C10orf32*, residing at the same locus as the *CYP17A1* gene and identified in large BP GWAS studies.[26,28] The *CYP17A1* gene has been identified as involved in a monogenic BP syndrome as well,[168] with variants leading to 17 alpha-hydroxylase deficiency resulting in congenital adrenal hyperplasia, a hallmark of which is hypertension.[169–171] It has also been assumed to be the gene of interest at the GWAS locus[166] because of its prior functional evidence, but our results show that, while *CYP17A1* is important for monogenic forms of hypertension, *C10orf32* may be the gene playing a role in essential hypertension, thereby, via a different mechanism relating to the vasculature. Additionally, despite specific evidence ruling out *CYP17A1* and supporting *C10orf32* in the arteries, other genes at this locus must be considered for future studies even in these tissues, because of the not infrequent occurrence of eQTLs affecting multiple neighboring genes in the same tissue, as seen in Chapter 4 (such as, the *MTHFR*/*CLCN6* and *CERS5*/*COX14*/*RP4-605O3.4* loci). This type of analysis, however, provides a valuable starting place to determine genes and variants of interest that can subsequently be studied

in greater depth for their functional effects in putatively relevant cell types. Furthermore, though the results of the enrichment analysis indicate the arteries as being of interest, the adrenal gland is another organ known to be involved in BP regulation.[172] Further study of this and other tissues or organs may reveal effects of genes specific to these sites of action.

An additional point of interest is the concurrent support of the *ULK4* gene associated with DBP from two gene discovery studies in different populations (Chapters 2 and 4). Though the Illumina HumanExome Beadchip on which the African-ancestry individuals were genotyped focused on rare and low-frequency coding variants as described in Chapter 2, a small subset of common variants represented on the chip replicated several previously associated common variants at this locus. Additionally, we found strong evidence for a low frequency coding variant supporting the gene's association with DBP. The analysis in Chapter 4 provided support for the association of this gene to DBP in European-ancestry individuals by linking its putative regulatory variants to the phenotype and to its expression in arteries.

The studies in Chapters 2 and 3 focused primarily on identification of rare variants with large effects for blood pressure and hypertension, which is a common disease. The Chapter 3 study notably focused on specific variants involved in rare syndromes resulting in hypotension (Gitelman and Bartter syndromes), which extends the applicability of the analysis and identification of rare variants with larger effects even to rare diseases. The analyses in Chapter 4 included variants across the allele frequency spectrum, and illustrated, in conjunction with the rare variant analyses, that defining pathogenicity of a variant must be done both at the level of annotation in the genome, as

well as in the context of specific and relevant tissues and cell-types. This is an important and significant conclusion from this dissertation.

More generally, because hypertension is the major risk factor for all cardiovascular disease,[1] discovering the genetic causes of BP regulation by extension informs us about cardiovascular disease genetics as a whole. The importance of the large arteries as highlighted in Chapter 4, particularly of specific arterial cell types, needs to be further studied to elucidate specific mechanisms of the genes of interest to hypertension, and in general, cardiovascular diseases.

Progress in our understanding of the involvement of kidney in hypertension to an even greater extent is essential. While kidney BP control systems, including the renin-angiotensin-aldosterone-system (RAAS) have been established, recent evidence for "non-classical" RAAS pathways has suggested novel players in the BP control processes and in kidney function.[173] Furthermore, hypertension is associated with chronic kidney disease,[1] and although the condition is referred to as "hypertension-attributed" chronic kidney disease or nephropathy, whether hypertension is primary or secondary to associated kidney disease has been often debated.[174,175] Additional study to answer this question may also reveal additional pathways influencing BP control.

When the clinical utilities of gene discovery are of interest, they normally involve identifying novel drug targets or therapies. While numerous anti-hypertensive therapies have been developed[2], nearly half of affected individuals remain insufficiently treated[3] and one possible reason may be the current incomplete understanding of BP genetics and pathways. Thus, elucidating the pathways and mechanisms in the appropriate tissues or organs allows the possibility of identifying an even larger range of specific therapies.

A number of other scientific advances will need to be made in order to further our understanding of BP genetics. In addition to our findings described in Chapter 4,[33] other recent findings analyzing BP GWAS results also implicate the endothelial/vasculature in BP regulation.[22,36] The availability of open chromatin and expression data across a variety of tissues and cell types have made analyses like this possible; however, the paucity of kidney data has rendered such studies for BP incomplete, and its availability would improve assessments of contributions of different tissues to BP regulation. Additionally, the availability of genotypes, expression, and phenotypes *for the same samples* will allow the integration and evaluation of these variables and lead to more robust inferences than possible with data from different studies, as was done in Chapter 4 with the GERA and GTEx studies. Finally, the increasing sample sizes of BP GWAS are expected to continue to identify even more novel loci associated with BP,[176] and the expansion in sample sizes would allow increased power to identify variants across the allele frequency spectrum.

In this dissertation, we identified rare, putatively functional protein-coding variants, which has suggested specific novel genes of interest; by incorporating tissue-specific information, we not only identified additional genes, but also information about the relevance of their expression to BP regulation, and in some cases, the specificity of their expression in relation to BP regulation. Given the primary results of the studies detailed in this dissertation, expanding our studies in Chapter 4 to other tissues and cell types would be of great interest and importance. In addition to the adrenal gland, as well as specific arterial cell types, studying open chromatin and expression data from a variety of tissues and cell types would also be informative in discerning the specificity of the

phenotype. As more functional genomic data become available, analyses will evolve to incorporate functional information directly into gene discovery with information about functional context, and will facilitate the construction of tissue- or even cell-type-specific regulatory networks. Further, evidence may more easily be combined from analyses of coding variants and regulatory variants to identify specific disease genes. This is necessary to narrow down the inevitably large lists of candidate genes and variants we currently generate, for feasibility of functional testing.

Appendix A: MD-GEM Wolfe Street Competition Publication

The publication resulting from a collaborative project funded by the Wolfe Street Competition (http://www.hopkinsgenetics.org/wsc/2014/) in 2014 follows. This version is the main text from the accepted preprint (PLOS ONE, open access publication).

The citation for the full article is:
Nandakumar P*, Tin A*, Grove ML, Ma J, Boerwinkle E, Coresh J, et al. (2017) MicroRNAs in the miR-17 and miR-15 families are downregulated in chronic kidney disease with hypertension. PLoS ONE 12(8): e0176734.
https://doi.org/10.1371/journal.pone.0176734

*denotes equal contributions

MicroRNAs in the miR-17 and miR-15 families are downregulated in chronic kidney disease with hypertension

Priyanka Nandakumar[1,2¶], Adrienne Tin[3¶], Megan L. Grove[4], Jianzhong Ma[4], Eric Boerwinkle[4], Josef Coresh[3], Aravinda Chakravarti[1*]

[1] McKusick - Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, United States of America

[2] Predoctoral Training Program in Human Genetics and Molecular Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America

[3] Johns Hopkins Bloomberg School of Public Health; Welch Center for Prevention, Epidemiology, and Clinical Research, Baltimore, Maryland, United States of America

[4] Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, United States of America

[*] Corresponding Author:
E-mail: aravinda@jhmi.edu

[¶]These authors contributed equally to this work.

Abstract

Background. In older adults (aged 70-74 years), African-Americans have 4-fold higher risk of developing hypertension-attributed end-stage renal disease (ESRD) than European-Americans. A hypothesized mechanism linking hypertension and progressive chronic kidney disease (CKD) is the innate immune response and inflammation. Persons with CKD are also more susceptible to infection. Gene expression in peripheral blood can provide a view of the innate immune activation profile. We aimed to identify differentially expressed genes, microRNAs, and pathways in peripheral blood between cases with CKD and high blood pressure under hypertension treatment versus controls without CKD and with controlled blood pressure in African Americans.

Methods. Case and control pairs (N=15x2) were selected from those without diabetes and were matched for age, sex, body mass index, *APOL1* risk allele count, and hypertension medication use. High blood pressure under hypertension treatment was defined as hypertension medication use and systolic blood pressure (SBP) ≥ 145 mmHg. CKD was defined as estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73m^2. Cases were selected from those with CKD and high blood pressure under hypertension treatment, and controls were selected from those without CKD (eGFR: 75-120 mL/min/1.73m^2 and urine albumin-to-creatinine ratio < 30mg/g) and with blood pressure controlled by hypertension medication use (SBP < 135 mmHg and D(diastolic)BP < 90 mm Hg). We perform RNA sequencing of mRNA and microRNA and conducted differential expression and co-expression network analysis.

Results. Of 347 miRNAs included in the analysis, 14 were significantly associated with case status (Benjamini-Hochberg adjusted p-value [BH p] < 0.05). Of these, ten were downregulated in cases: three of each belong to the miR-17 and miR-15 families. In co-expression network analysis of miRNA, one module, which included 13 of the 14 significant miRNAs, had significant association with case status. Of the 14,488 genes and 41,739 transcripts included in the analysis, none had significant association with case status. Gene co-expression network analyses did not yield any significant associations for mRNA.

Conclusion. We have identified 14 differentially expressed miRNAs in the peripheral blood of CKD cases with high blood pressure under hypertension treatment as compared to appropriate controls. Most of the significant miRNAs were downregulated and have critical roles in immune cell functions. Future studies are needed to replicate our findings and determine whether the downregulation of these miRNAs in immune cells may influence CKD progression or complications.

Introduction

About one-third of end-stage renal disease (ESRD) in the U.S. has been attributed to hypertension [1,2]. In older adults, African-Americans have 4-fold higher risk of developing hypertension-attributed ESRD than European-Americans, and the incidence of hypertension-attributed ESRD increases with age [3]. A hypothesized mechanism linking hypertension and kidney function decline is the innate immune response and inflammation [4]. Inflammation biomarkers in blood have been associated with kidney function decline and incident hypertension [5-8]. Additionally, older adults with chronic kidney disease (CKD) have increased risk of infection-related hospitalization or bloodstream infection [9,10]. Although immune dysfunction in persons with ESRD due to metabolic disorder and retention of uremic solute is well established [11,12], the mechanisms underlying the links between immune function, hypertension, and kidney function in earlier stages of CKD is not well understood. Gene expression analysis in peripheral blood can provide a view of the innate immune activation profile and may lead to insights into the pathophysiology of hypertension and CKD, and their complications.

Studies using immune cells have identified specific gene expression profiles of CKD patients versus hemodialyzed patients, and of patients with essential hypertension versus controls [13,14]. Recently microRNAs (miRNA) have been shown to have a potential role in influencing blood pressure and kidney function through modulating the immune response [15-17]. miRNAs are small noncoding RNAs (~22 nucleotides in length) with important post-transcriptional regulatory functions and are an attractive target of investigations.

Given that African Americans have a strong predisposition for hypertension-attributed kidney disease, we decided to study hypertension and kidney disease jointly in older African Americans. We aimed to identify differentially expressed genes and miRNAs in the peripheral blood to gain insights into the immune profile of hypertension and kidney disease. We used a matched case-control design (n=15x2) in the African Americans cohort of the Atherosclerosis Risk in Communities (ARIC) study [18]. We conducted RNA sequencing to quantify expression levels and performed differential expression and co-expression network analyses to identify gene expression profiles in the peripheral blood of hypertension and CKD cases.

Methods

Ethics statement

Written informed consent was obtained from all ARIC study participants, and approval was granted from the relevant institutional review boards (IRB) for the participating study centers (University of North Carolina, University of Minnesota, University of Mississippi Medical Center, and Johns Hopkins University). JHSPH IRB No. H.34.99.07.02.A1, study title "Atherosclerosis Risk in Communities (ARIC) Study - Morbidity/Mortality Follow-up Field Center." This research was conducted in accordance with the principles described in the Declaration of Helsinki.

Study design

The ARIC Study is an ongoing prospective cohort study in four US communities [18]. A total of 15,792 participants aged 45–64 years were recruited from Forsyth County, North Carolina; Jackson, Mississippi; suburban Minneapolis, Minnesota; and Washington County, Maryland between 1987 and 1989 (Visit 1). Four follow-up examinations (visits 2–5) have been conducted. Blood samples preserved for RNA analysis using PAXgene blood tubes were available from visit 5 (2011-13).

Case and control pairs (N=15 pairs) were selected from those individuals without diabetes but on hypertension treatment and they were matched by age, sex, body mass index, *APOL1* renal risk allele count, and hypertension medications (ACE inhibitor, angiotensin receptor blocker, and calcium channel blocker) to reduce heterogeneity. The use of hypertension medication was included in the selection criteria because a high proportion (>60%) of older African Americans were on hypertension treatment [19], and some hypertension medications have been reported to influence gene expression in peripheral blood [20]. In the ARIC study, hypertension medication use was determined by inspection of medication bottles at study visit. High blood pressure under hypertension treatment was defined as systolic blood pressure (SBP) ≥ 145 mmHg with hypertension medication use. CKD was defined as estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73m^2. Cases were selected from those with CKD and high blood pressure under hypertension treatment. Controls were selected from those with blood pressure controlled by hypertensive medications (SBP < 135 mmHg and D(diastolic)BP < 90 mm Hg) and without CKD (eGFR: 75-120 mL/min/1.73m2 and urine albumin-to-creatinine ratio [UACR] < 30mg/g). The *APOL1* renal risk variants are strongly associated with CKD progression [21], thus matching by *APOL1* risk variants provided the opportunity to identify differential expressed genes that were independent of *APOL1*. The characteristics of the cases and controls were compared using the t-test, Kruskal-Wallis, or Fisher's exact test, as appropriate.

mRNA and miRNA library preparation, sequencing, processing, alignment, and quantitation methods are described in S1 Text.

Surrogate variable generation

Surrogate variables (SVs) representing unknown confounders were estimated using the svaseq function from the R package sva [22], with the null model absent of variables, and case-status as the variable of interest in the full model. SVs were included as covariates in models in all subsequent expression analyses as expanded upon in the Differential Expression Analyses Methods sections.

mRNA differential expression analyses

The mRNA analyses were conducted on genes and transcripts from a list of 20,377 genes tagged as "protein-coding" in the GENCODE V19 annotation files. Results were adjusted within each analysis using the Benjamini-Hochberg [23] procedure for multiple testing.

The R package DESeq2 [24] was used to conduct gene-level analyses, with the null hypothesis of a zero log2-fold change. As DESeq2 incorporates normalization and default outlier replacement procedures, SVs were generated on DESeq2 outlier-replaced, normalized counts and included as covariates for these analyses. Gene-level analyses were carried out on14,488 genes with a normalized count ≥1 in at least 14 out of 29

samples (~50% as one sample failed library preparation, see Results), with a target FDR of alpha=0.05 used for independent filtering. SVs were generated using 10,524 genes with a minimum normalized count of 10 in all 29 samples.

Gene-level kidney-focused analyses were also conducted, in which 397 kidney-expressed proteins from the Human Protein Atlas [25] (v14) (http://www.proteinatlas.org/humanproteome/kidney) were examined. Of these, 392 were available in our data; four were not available in the ENSEMBL GRCh37 assembly, and one was not in our final GTF of protein-coding genes. Genes meeting the normalized count threshold defined as above for expression were subset to produce the final kidney-specific results.

The Ballgown [26] R package was used to analyze 41,739 transcripts with fragments per kilobase of transcript per million mapped reads (FPKM) ≥ 0.3 in at least 14 samples, and SVs were generated on transcripts with FPKM ≥ 0.5 in at least 26 samples (90%) (17,085 transcripts).

miRNA differential expression analyses

The miRNA analyses were conducted on mature miRNAs from miRBase [27] v20 and novel miRNAs, detected as described in S1 Text. Similar to the mRNA analyses, the R package DESeq2 was used for miRNA analyses with the null hypothesis of a zero log2-fold change, and SVs were generated on the outlier-replaced normalized counts and included as covariates in the model. These analyses were carried out on 347 miRNAs with a normalized count ≥ 1 in at least 15 out of 30 samples (50%) with target FDR of alpha=0.05 used for independent filtering, and SVs were generated using 270 miRNAs with a minimum normalized count of 3 in at least 27 samples (90%). Counts were summed across all precursors for each mature miRNA. Results were adjusted using the Benjamini-Hochberg procedure for multiple testing. We performed hierarchical clustering of the normalized counts of the miRNAs with nominal p-value < 0.01 to visualize the expression levels of these miRNAs between cases and controls.

Co-expression network analysis

We used the Weighted Gene Co-Expression Network Analysis (WGCNA) R package to construct correlated signed network and tested for associations between the eigengene (the first principal component) of each module and case status [28]. The genes, transcripts, and miRNAs used for co-expression network analysis and the generation of surrogate variables was the same as those for differential expression analyses. The gene and miRNA counts were first normalized by the size factor with outlier replacement using DESeq2 [24]. The transcript counts were normalized to FPKM using Ballgown [26]. Next, the gene, transcript, or miRNA expression levels were transformed using natural logs after adding 1. Residuals of the log transformed expression levels were generated with adjustment for surrogate variables. These residuals were used as input for co-expression analysis. In network construction, we used bi-weight mid-correlation as a measure of co-expression to minimize the influence from outliers. For soft thresholding power, we used the first power with adjusted R square for linear fit > 0.8. The minimum module size was set as 25, 30, and 8 for genes, transcripts, and miRNA, respectively. The significant threshold for module association was defined as 0.05 divided by the number of correlated modules in each specific analysis.

miRNA target prediction and gene ontology analyses

The procedure for miRNA target prediction and gene ontology analyses of differentially expressed miRNAs is similar to the procedure described by the authors of the empiricalGO [29] software in Bleazard et al. 2015. Briefly, miRanda [30] v3.3a was used for miRNA target prediction with parameters free energy < −20 kcal/mol and score > 155 [29]. The 3'UTR sequences for target prediction were obtained from GRCh37 ENSEMBL BioMart [31]. Subsequently, gene ontology (GO) annotations were obtained from GRCh37 ENSEMBL BioMart, and the empiricalGO python script was used in "basic" mode to produce a list of GO terms with a one-tailed permutated p-value for each term. The final results were subset to include only terms with a minimum size of 10, and the Benjamini-Hochberg multiple testing correction was applied to adjust for the number of GO terms in each list.

Results

Study population characteristics

The mean age of the cases and controls was 77 years, and 67% were female. No significant differences were detected between cases and controls for all of the matching characteristics (**Table 1)**. The cases had significantly lower eGFR (mean of 46 min/mL/1.73m^2 vs 88 min/mL/1.73m^2, $p < 0.0001$) and higher SBP (mean of 156 mm Hg vs. 115 mm Hg, $p < 0.0001$). UACR was higher in cases (median of 13.7mg/g vs 7.8mg/g), but the difference was not significant (p=0.13). On hypertension medications, none of the participants were on both angiotensin-converting-enzyme (ACE) inhibitor and angiotensin receptor blocker while two participants were on ACE inhibitor and beta blocker, and four participants were on ACE inhibitor and calcium channel blocker. These participants were split evenly between the case and control groups. Since gene expression levels were measured in whole blood, we also compared cases and controls for white blood cell count and percentage of lymphocyte, monocyte, and granulocyte, and did not observe significant differences.

Table 1. Study population characteristics

Variable	Case	Control	p
Selection variables			
eGFR, mL/min/m^2, mean (SD)	46 (12)	88 (10)	N/A
SBP, mm Hg, mean (SD)	156 (12)	115 (33)	N/A
Matching variables			
Age, year, mean (SD)	77 (4.5)	77 (5.4)	0.97
Female, % (n)	67 (10)	67 (10)	1.00
BMI, kg/m^2, mean (SD)	28 (3.5)	30 (3.3)	0.09
APOL1 high-risk, % (n)	33 (5)	33 (5)	1.00
Hypertension medications			
ACE inhibitor, % (n)	53.3 (8)	60 (9)	1.00
Angiotensin receptor blocker, % (n)	27 (4)	27 (4)	1.00

Beta blocker, % (n)	33.3 (5)	33.3 (5)	1.00
Calcium channel blocker, % (n)	47 (7)	60 (9)	0.71
Blood cell type variables			
White blood cell count, 1000 per $mm^{3,}$ median (1st, 3rd quartile)	5.4 (4.1, 6.2)	4.8 (4.4, 5.9)	0.80
Lymphocyte, %, median (1st, 3rd quartile)	33.6 (29.1, 39.0)	30.4 (27.8, 40)	0.72
Monocyte, %, median (1st, 3rd quartile)	13.8 (12.8, 17.2)	13.4 (12.5, 16.0)	0.48
Granulocyte, %, median (1st, 3rd quartile)	50.9 (45.1, 57.4)	54.8 (46.3, 59.8)	0.48
Other variables			
DBP, mm Hg, mean (SD)	76 (12)	70 (19)	0.31
Albuminuria, median (1st, 3rd quartile)	13.7 (5.9, 66.7)	7.8 (5.7, 13.3)	0.13
High sensitive C-reactive Protein,	1.4 (1.01, 6.87)	5.3 (2.5, 6.3)	0.31
Serum creatinine, mg/dL	1.6 (0.9)	0.8 (0.1)	0.003
Diuretic use, % (n)	73.3 (11)	64.3 (9)	0.70

p, p-value, N/A, not applicable

Sequencing and processing of mRNAs and miRNAs

Sequencing of mRNA was successful for 29 out of 30 samples; one sample (case) failed library preparation. All 30 samples were successfully sequenced for miRNA. Quality control (QC) and mapping statistics for the 29 samples with mRNA data (depth: 18.7M-45.1M paired-end reads) and the 30 samples with miRNA data (depth: 6.2M-11.5M single-end reads) are listed in S1 and S2 Tables, respectively. As S1 Table indicates, External RNA Controls Consortium (ERCC) transcripts were poorly detected in one sample, but this individual was retained for analysis because the distribution of counts in other genomic features aligned with those of the other samples, indicating a possible issue with the ERCC spike-in itself.

Differential expression analysis of genes, transcripts, and miRNAs

The gene-level differential expression analysis of 14,448 protein-coding genes considered "expressed" (normalized count ≥ 1 in ≥50% of samples) produced no significant genome-wide results after Benjamini-Hochberg (BH) multiple testing correction (S3 Table). The relevance of the kidney to both hypertension and CKD led us to examine a set of 154 genes meeting our expression threshold and expressed in the kidney, based on data available from the Human Protein Atlas [25], which contains protein expression data from four kidney samples, with proteins in the glomeruli, proximal tubules, distal tubules and the collecting ducts. In this subset of genes, none were significant after BH correction. The gene with the lowest p-value was *SMIM24* (unadjusted $p=5.67x10^{-4}$, BH p=0.087).

Transcript-level analyses were conducted on 41,739 transcripts with FPKM ≥ 0.3 in ≥ 50% of samples. After BH correction, there were no differentially expressed transcripts identified; results with $p < 10^{-3}$ are presented in S4 Table.

We tested 347 miRNAs, including 12 novel miRNAs, for differential expression. We detected 14 significant miRNAs. Four were upregulated, and 10 were downregulated in the cases as compared to the controls (Table 2). The most significant miRNA was miR-17-5p (log2 fold change=-0.77, BH p=6.7E-4). Two other miRNAs in the miR-17 family (miR-106a-5p, miR-106b-3p) were also significantly downregulated. The miR-15 family has three members that were significantly downregulated in the cases (miR-15a-5p, miR-15b-5p, and miR-16-5p). Twenty members of the miR-17 and miR-15 families were found in our data and all, but two, were downregulated in cases, although the associations of some were not significant (S5 Table). S1 Fig presents a heat map of the normalized count of the 32 miRNAs with nominal p-value < 0.01.

Table 2. Differentially expressed miRNAs (BH p<0.05)

miRNA	Mean of normalized count	Log2 Fold Change	SE	p	BH p
hsa-miR-17-5p	71.97	-0.77	0.16	1.94×10^{-6}	6.76×10^{-4}
hsa-miR-130a-3p	492.19	-0.60	0.15	4.31×10^{-5}	7.47×10^{-3}
hsa-miR-15b-5p	215.32	-0.72	0.19	1.24×10^{-4}	1.17×10^{-2}
hsa-miR-106b-3p	1425.98	-0.65	0.17	1.35×10^{-4}	1.17×10^{-2}
hsa-miR-106a-5p	5.46	-1.03	0.28	2.12×10^{-4}	1.34×10^{-2}
hsa-miR-16-5p	6983.54	-0.64	0.17	2.33×10^{-4}	1.34×10^{-2}
hsa-miR-181a-5p	15301.70	-0.59	0.16	2.80×10^{-4}	1.39×10^{-2}
hsa-miR-1285-3p	561.01	-0.44	0.13	4.92×10^{-4}	2.14×10^{-2}
hsa-miR-15a-5p	430.52	-0.76	0.23	8.36×10^{-4}	3.18×10^{-2}
hsa-miR-29c-5p	73.85	0.55	0.17	9.18×10^{-4}	3.18×10^{-2}
hsa-miR-345-5p	715.07	0.60	0.18	1.04×10^{-3}	3.27×10^{-2}
hsa-miR-142-3p	89.46	0.65	0.20	1.23×10^{-3}	3.52×10^{-2}
hsa-miR-339-3p	507.26	0.36	0.11	1.32×10^{-3}	3.52×10^{-2}
hsa-miR-210-3p	440.49	-0.39	0.12	1.50×10^{-3}	3.72×10^{-2}

SE, standard error; p, p-value; BH p, Benjamini-Hochberg adjusted p-value

Experimentally validated targets in humans were available for 11 of the 14 significant miRNAs in miRTarBase [32] 6.0 (S6 Table). The number of validated target genes for each significant miRNA ranged from 4 (miR-1285-3p) to 134 (miR-17-5p), with a median of 56. Altogether 272 unique genes have experimental evidence of being regulated by the 11 significant miRNAs. Of these, 239 were detected in our data, and 8 were associated with case status at p < 0.05 (S7 Table).

Co-expression network analysis of genes and transcripts

To investigate whether patterns in gene co-expression may be related to case-status, we conducted gene co-expression analysis. Of the 14,488 genes included in the analysis, 178 were assigned into four co-expression modules (S8 Table). The rest were pruned by WGCNA due to low correlation (< 0.3) with the eigengene in each module.

The association between the eigengene of the brown module (consisting of 35 genes) and case status was nominally significant (p=0.03, S9 Table). Of the 41,739 transcripts included in the analysis, no correlated modules were detected.

Co-expression network analysis of miRNAs

Of the 347 miRNA included in co-expression analysis, 182 were assigned into five co-expression modules with correlated miRNAs (S10 Table). The rest (n=165) were pruned due to low correlation with the eigengene in each module. The eigengene of the turquoise module (108 miRNAs) was significantly associated with case status (p=0.005, Table 3). The eigengene of the blue module (29 miRNAs) was nominally associated with case status (p=0.03). For both modules, case status was associated with lower eigengene values suggesting an association with lower miRNA expression in these modules. The turquoise module included 13 of the 14 miRNAs that were significantly downregulated in the cases (Table 2).

Table 3. Association between the eigengene of each miRNA co-expression module and case status

Module	Beta[b]	SE	p[a]	Number of miRNAs
Turquoise[c]	-0.97	0.32	$5.38x10^{-3}$	108
Blue	-0.80	0.34	$2.63x10^{-2}$	29
Green	-0.06	0.37	$8.63x10^{-1}$	11
Brown	0.03	0.37	$9.32x10^{-1}$	23
Yellow	-0.01	0.37	$9.84x10^{-1}$	11

SE, standard error; p, p-value;

[b] Beta values are in log expression level units.

[c] 13 out of 14 differentially expressed miRNAs (excluding hsa-miR-339-3p), belong to the turquoise module.

Gene ontology analysis of significant miRNAs

We then proceeded to predict target genes of the significant miRNAs using miRanda. The predicted targets were analyzed for enrichment of gene ontology annotations using empiricalGO with all predicted targets of the 347 expressed miRNAs as the "universe." empiricalGO counted 5,576 targets for the 14 differentially expressed miRNAs from the target genes, of which 4,431 had associated GO terms. No GO terms with minimum size of 10 genes in our data were significant after Benjamini-Hochberg multiple test correction, while 25 terms had a one-tailed permutated $p < 0.05$ (S11 Table). All results for mRNA and miRNA analyses are summarized in S2 Fig.

Discussion

Main findings

We have identified 14 miRNAs that were significantly associated with CKD and high blood pressure under hypertension treatment. Ten of these miRNAs were downregulated in the cases, and 13 were grouped in one module in co-expression

network analysis. Three of the 14 miRNAs belong to the miR-17 family, and three belong to the miR-15 family.

The miRNAs that were significantly associated with case status have critical roles in immune cell function. First, miR-17-5p, the most significantly downregulated miRNA, belongs to a cluster of miRNAs located in intron 3 of *C13orf25* at chromosome 13 [33]. This cluster of miRNAs (miR-17/92) has a wide range of functions in immune cell development and differentiation [33]. Specifically, the miR-17 cluster has been found to promote T cell survival, regulate Th1 response, and interleukin 10 (IL10) production in regulatory T cells [34,35]. Members of two other clusters (miR-106a/363 and miR-106b/25) of the miR-17 family were also significantly downregulated in cases. miR-106a on the X chromosome has been shown to downregulate IL10 expression [36], and all three miR-17 family members differentially expressed in our study (106a, 106b, 17-5p) are known to be upregulated in activated T lymphocytes [37]. Second, three members of the miR-15 family (miR-15a-5p, miR-15b-5p, miR-16-5p) were also significantly downregulated in cases. Expression of miR-15 has been found to enhance the induction of regulatory T cells from naïve CD4+ T cells [38]. Additionally, miR-210 suppresses proinflammatory cytokines in murine macrophages [39]. Finally, in considering experimentally verified targets of the differentially expressed miRNAs in this study, the *BCL2*, *CCND1*, and *VEGFA* genes are each targeted by five miRNAs from at least two miRNA families. These genes are known factors in apoptosis and cell survival, having previously been studied in the context of cancer pathways [40,41]. Taken together, the downregulation of the above miRNAs in cases suggests a lower immune activation state. Whether this lower activation state has implications for CKD progression or complications requires further investigation.

Population-based studies have reported increased risk of infection among persons with early stages of CKD. In older adults (age $\geq$ 65 years), reduced kidney function (eGFR < 60 mL/min/1.73m^2) is associated with increased risk of infection-related hospitalization or bloodstream infection [9,10]. In persons with ESRD, immune dysfunction due to metabolic disorder and retention of uremic solute is well established [11,12]. The decline in kidney function occurs in a continuum. Thus, immune dysfunction may play a role in earlier stages of CKD. The increased risk of infection in persons with earlier stages of CKD is consistent with our findings of the downregulation of some miRNAs that are critical for immune cell function.

On kidney disease, the plasma levels of miR-15b were found to be 2-fold lower in hemodialysis patients versus non-CKD controls [42]. This result is consistent with our study. The associations between levels of miR-15b and miR-17 in kidney tissues and acute kidney injury in human and in animal models have been reported [43]. Since gene expression levels are highly tissue specific, it is uncertain whether studies of gene expression in one tissue may be generalizable to a different tissue [44-48].

Strengths and limitations

One of the strengths of this study is its design. The cases and controls were carefully selected and matched for important potential confounders of gene expression, including hypertension medication use. In addition, we performed deep sequencing of the miRNA pool that allowed the discovery of novel miRNAs. However, our study has some limitations as well. First, the differentially expressed miRNAs need to be replicated in an

independent study and validated by alternate laboratory methods, such as real-time polymerase chain reaction. Second, gene expression levels were measured in whole blood and not in specific types of immune cells. Thus, although cases and controls did not differ in major white blood cell types, and we used surrogate variables to control for unmeasured confounders, we cannot exclude the possibility that the differentially expressed miRNAs arise from differences in the distributions of white blood cell subpopulations. Third, our study is a cross-sectional study. We cannot distinguish whether the differential expression of miRNAs might have influenced disease development or was a consequence of the disease condition. Fourth, in contrast to miRNA co-expression network analysis, mRNA co-expression network analysis did not detect any significant modules although one module was nominally significant. This suggests larger sample size may be required for mRNA differential expression analysis. Finally, our cases and controls were under hypertension treatment, therefore our results cannot be generalized to persons without hypertension. Since the majority (>60%) of older adults in the U.S. are hypertensive [19], our results are, however, relevant to a large proportion of older adults.

Acknowledgements
The authors thank the staff and participants of the ARIC study for their important contributions.

References

References
1. U.S. Renal Data System, USRDS 2012 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2012.
2. The data reported here have been supplied by the United States Renal Data System (USRDS). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.
3. Kopp JB. Rethinking hypertensive kidney disease: arterionephrosclerosis as a genetic, metabolic, and inflammatory disorder. Curr Opin Nephrol Hypertens. 2013 May;22(3):266–72.
4. Harrison DG, Guzik TJ, Lob HE, Madhur MS, Marvar PJ, Thabet SR, et al. Inflammation, immunity, and hypertension. Hypertens Dallas Tex 1979. 2011 Feb;57(2):132–40.
5. Erlinger TP, Tarver-Carr ME, Powe NR, Appel LJ, Coresh J, Eberhardt MS, et al. Leukocytosis, hypoalbuminemia, and the risk for chronic kidney disease in US adults. Am J Kidney Dis Off J Natl Kidney Found. 2003 Aug;42(2):256–63.
6. Bash LD, Erlinger TP, Coresh J, Marsh-Manzi J, Folsom AR, Astor BC. Inflammation, hemostasis, and the risk of kidney function decline in the Atherosclerosis Risk in Communities (ARIC) Study. Am J Kidney Dis Off J Natl Kidney Found. 2009 Apr;53(4):596–605.

7. Sesso HD, Jiménez MC, Wang L, Ridker PM, Buring JE, Gaziano JM. Plasma Inflammatory Markers and the Risk of Developing Hypertension in Men. J Am Heart Assoc. 2015 Sep;4(9):e001802.
8. Mattace-Raso FUS, Verwoert GC, Hofman A, Witteman JCM. Inflammation and incident-isolated systolic hypertension in older adults: the Rotterdam study. J Hypertens. 2010 May;28(5):892–5.
9. Dalrymple LS, Katz R, Kestenbaum B, de Boer IH, Fried L, Sarnak MJ, et al. The risk of infection-related hospitalization with decreased kidney function. Am J Kidney Dis Off J Natl Kidney Found. 2012 Mar;59(3):356–63.
10. James MT, Laupland KB, Tonelli M, Manns BJ, Culleton BF, Hemmelgarn BR, et al. Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis. Arch Intern Med. 2008 Nov 24;168(21):2333–9.
11. Vanholder R, Ringoir S. Infectious morbidity and defects of phagocytic function in end-stage renal disease: a review. J Am Soc Nephrol JASN. 1993 Mar;3(9):1541–54.
12. Chonchol M. Neutrophil dysfunction and infection risk in end-stage renal disease. Semin Dial. 2006 Aug;19(4):291–6.
13. Zaza G, Granata S, Rascio F, Pontrelli P, Dell'Oglio MP, Cox SN, et al. A specific immune transcriptomic profile discriminates chronic kidney disease patients in predialysis from hemodialyzed patients. BMC Med Genomics. 2013;6:17.
14. Marketou ME, Kontaraki J, Zacharis E, Parthenakis F, Maragkoudakis S, Gavras I, et al. Differential gene expression of bradykinin receptors 1 and 2 in peripheral monocytes from patients with essential hypertension. J Hum Hypertens. 2014 Jul;28(7):450–5.
15. Trionfini P, Benigni A, Remuzzi G. MicroRNAs in kidney physiology and disease. Nat Rev Nephrol. 2015 Jan;11(1):23–33.
16. Ichii O, Otsuka S, Sasaki N, Namiki Y, Hashimoto Y, Kon Y. Altered expression of microRNA miR-146a correlates with the development of chronic renal inflammation. Kidney Int. 2012 Feb;81(3):280–92.
17. Li S, Zhu J, Zhang W, Chen Y, Zhang K, Popescu LM, et al. Signature microRNA expression profile of essential hypertension and its novel link to human cytomegalovirus infection. Circulation. 2011 Jul 12;124(2):175–84.
18. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989 Apr;129(4):687–702.
19. Egan BM, Zhao Y, Axon RN. US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008. JAMA. 2010 May 26;303(20):2043–50.
20. Napoleone E, Di Santo A, Camera M, Tremoli E, Lorenzet R. Angiotensin-converting enzyme inhibitors downregulate tissue factor synthesis in monocytes. Circ Res. 2000 Feb 4;86(2):139–43.
21. Parsa A, Kao WHL, Xie D, Astor BC, Li M, Hsu C, et al. APOL1 risk variants, race, and progression of chronic kidney disease. N Engl J Med. 2013 Dec 5;369(23):2183–96.
22. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. 2014 Dec 1;42(21).
23. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.

24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
25. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015 Jan 23;347(6220):1260419.
26. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol. 2015 Mar;33(3):243–6.
27. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014 Jan;42(Database issue):D68–73.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
29. Bleazard T, Lamb JA, Griffiths-Jones S. Bias in microRNA functional enrichment analysis. Bioinforma Oxf Engl. 2015 May 15;31(10):1592–8.
30. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003;5(1):R1.
31. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database J Biol Databases Curation. 2011;2011:bar030.
32. Chou C-H, Chang N-W, Shrestha S, Hsu S-D, Lin Y-L, Lee W-H, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2016 Jan 4;44(D1):D239–47.
33. Mogilyansky E, Rigoutsos I. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. Cell Death Differ. 2013 Dec;20(12):1603–14.
34. Jiang S, Li C, Olive V, Lykken E, Feng F, Sevilla J, et al. Molecular dissection of the miR-17-92 cluster's critical dual roles in promoting Th1 responses and preventing inducible Treg differentiation. Blood. 2011 Nov 17;118(20):5487–97.
35. de Kouchkovsky D, Esensten JH, Rosenthal WL, Morar MM, Bluestone JA, Jeker LT. microRNA-17-92 regulates IL-10 production by regulatory T cells and control of experimental autoimmune encephalomyelitis. J Immunol Baltim Md 1950. 2013 Aug 15;191(4):1594–605.
36. Sharma A, Kumar M, Aich J, Hariharan M, Brahmachari SK, Agrawal A, et al. Posttranscriptional regulation of interleukin-10 expression by hsa-miR-106a. Proc Natl Acad Sci U S A. 2009 Apr 7;106(14):5761–6.
37. Grigoryev YA, Kurian SM, Hart T, Nakorchevsky AA, Chen C, Campbell D, et al. MicroRNA regulation of molecular networks mapped by global microRNA, mRNA, and protein expression in activated T lymphocytes. J Immunol Baltim Md 1950. 2011 Sep 1;187(5):2233–43.
38. Singh Y, Garden OA, Lang F, Cobb BS. MicroRNA-15b/16 Enhances the Induction of Regulatory T Cells by Regulating the Expression of Rictor and mTOR. J Immunol Baltim Md 1950. 2015 Dec 15;195(12):5667–77.
39. Qi J, Qiao Y, Wang P, Li S, Zhao W, Gao C. microRNA-210 negatively regulates LPS-induced production of proinflammatory cytokines by targeting NF-κB1 in murine macrophages. FEBS Lett. 2012 Apr 24;586(8):1201–7.

40. Tang S-W, Chang W-H, Su Y-C, Chen Y-C, Lai Y-H, Wu P-T, et al. MYC pathway is activated in clear cell renal cell carcinoma and essential for proliferation of clear cell renal cell carcinoma cells. Cancer Lett. 2009 Jan 8;273(1):35–43.
41. Wang H, Xu Z, Ma M, Wang N, Wang K. Network analysis of microRNAs, transcription factors, target genes and host genes in nasopharyngeal carcinoma. Oncol Lett. 2016 Jun;11(6):3821–8.
42. Wang H, Peng W, Ouyang X, Dai Y. Reduced circulating miR-15b is correlated with phosphate metabolism in patients with end-stage renal disease on maintenance hemodialysis. Ren Fail. 2012;34(6):685–90.
43. Fan P-C, Chen C-C, Chen Y-C, Chang Y-S, Chu P-H. MicroRNAs in acute kidney injury. Hum Genomics. 2016 Sep 8;10(1):29.
44. Rodwell GEJ, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, et al. A transcriptional profile of aging in the human kidney. PLoS Biol. 2004 Dec;2(12):e427.
45. Fu J, Wolfs MGM, Deelen P, Westra H-J, Fehrmann RSN, Te Meerman GJ, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 2012 Jan;8(1):e1002431.
46. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015 May 8;348(6235):648–60.
47. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res. 2016 May 5;44(8):3865–77.
48. Fehlmann T, Ludwig N, Backes C, Meese E, Keller A. Distribution of microRNA biomarker candidates in solid tissues and body fluids. RNA Biol. 2016 Nov;13(11):1084–8.
49. Murtagh, F. & Legendre, P. J Classif (2014) 31: 274. doi:10.1007/s00357-014-9161-z.

Bibliography

1. Centers for Disease Control and Prevention (CDC). Vital signs: prevalence, treatment, and control of hypertension--United States, 1999-2002 and 2005-2008. *MMWR Morb Mortal Wkly Rep*. 2011;60(4):103-108.

2. Oparil S, Schmieder RE. New approaches in the treatment of hypertension. *Circ Res*. 2015;116(6):1074-1095. doi:10.1161/CIRCRESAHA.116.303603.

3. Yoon S, Fryar C, Carroll M. Hypertension Prevalence and Control Among Adults: United States, 2011–2014. NCHS Data Brief, No. 220. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and Prevention, US Dept of Health and Human Services, 2015.

4. Ehret GB. Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Curr Hypertens Rep*. 2010;12(1):17-25. doi:10.1007/s11906-009-0086-6.

5. Miall WE, Oldham PD. The hereditary factor in arterial blood-pressure. *Br Med J*. 1963;1(5323):75-80.

6. Salfati E, Morrison AC, Boerwinkle E, Chakravarti A. Direct Estimates of the Genomic Contributions to Blood Pressure Heritability within a Population-Based Cohort (ARIC). *PloS One*. 2015;10(7):e0133031. doi:10.1371/journal.pone.0133031.

7. Sanders PW. Dietary salt intake, salt sensitivity, and cardiovascular health. *Hypertens Dallas Tex 1979*. 2009;53(3):442-445. doi:10.1161/HYPERTENSIONAHA.108.120303.

8. Glover M, Zuber AM, O'Shaughnessy KM. Hypertension, dietary salt intake, and the role of the thiazide-sensitive sodium chloride transporter NCCT. *Cardiovasc Ther*. 2011;29(1):68-76. doi:10.1111/j.1755-5922.2010.00180.x.

9. Lifton RP, Gharavi AG, Geller DS. Molecular mechanisms of human hypertension. *Cell*. 2001;104(4):545-556.

10. Ji W, Foo JN, O'Roak BJ, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet*. 2008;40(5):592-599. doi:10.1038/ng.118.

11. Scholl UI, Goh G, Stölting G, et al. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nat Genet*. 2013;45(9):1050-1054. doi:10.1038/ng.2695.

12. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, et al. Genetic variants in novel pathways influence blood

pressure and cardiovascular disease risk. *Nature*. 2011;478(7367):103-109. doi:10.1038/nature10405.

13. Johnson T, Gaunt TR, Newhouse SJ, et al. Blood pressure loci identified with a gene-centric array. *Am J Hum Genet*. 2011;89(6):688-700. doi:10.1016/j.ajhg.2011.10.013.

14. Padmanabhan S, Melander O, Johnson T, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genet*. 2010;6(10):e1001177. doi:10.1371/journal.pgen.1001177.

15. Kato N, Takeuchi F, Tabara Y, et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat Genet*. 2011;43(6):531-538. doi:10.1038/ng.834.

16. Wain LV, Verwoert GC, O'Reilly PF, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet*. 2011;43(10):1005-1011. doi:10.1038/ng.922.

17. Franceschini N, Fox E, Zhang Z, et al. Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations. *Am J Hum Genet*. 2013;93(3):545-554. doi:10.1016/j.ajhg.2013.07.010.

18. Ganesh SK, Chasman DI, Larson MG, et al. Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations. *Am J Hum Genet*. 2014;95(1):49-65. doi:10.1016/j.ajhg.2014.06.002.

19. Tragante V, Barnes MR, Ganesh SK, et al. Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. *Am J Hum Genet*. 2014;94(3):349-360. doi:10.1016/j.ajhg.2013.12.016.

20. Simino J, Shi G, Bis JC, et al. Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia. *Am J Hum Genet*. 2014;95(1):24-38. doi:10.1016/j.ajhg.2014.05.010.

21. Wang Y, O'Connell JR, McArdle PF, et al. From the Cover: Whole-genome association study identifies STK39 as a hypertension susceptibility gene. *Proc Natl Acad Sci U S A*. 2009;106(1):226-231. doi:10.1073/pnas.0808358106.

22. Ehret GB, Ferreira T, Chasman DI, et al. The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat Genet*. 2016;48(10):1171-1184. doi:10.1038/ng.3667.

23. Johnson AD, Newton-Cheh C, Chasman DI, et al. Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals. *Hypertens*

Dallas Tex 1979. 2011;57(5):903-910. doi:10.1161/HYPERTENSIONAHA.110.158667.

24. Ganesh SK, Tragante V, Guo W, et al. Loci influencing blood pressure identified using a cardiovascular gene-centric array. *Hum Mol Genet*. 2013;22(8):1663-1678. doi:10.1093/hmg/dds555.

25. Takeuchi F, Isono M, Katsuya T, et al. Blood pressure and hypertension are associated with 7 loci in the Japanese population. *Circulation*. 2010;121(21):2302-2309. doi:10.1161/CIRCULATIONAHA.109.904664.

26. Levy D, Ehret GB, Rice K, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41(6):677-687. doi:10.1038/ng.384.

27. Newton-Cheh C, Larson MG, Vasan RS, et al. Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure. *Nat Genet*. 2009;41(3):348-353. doi:10.1038/ng.328.

28. Newton-Cheh C, Johnson T, Gateva V, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009;41(6):666-676. doi:10.1038/ng.361.

29. Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet*. 2015;47(11):1282-1293. doi:10.1038/ng.3405.

30. Parmar PG, Taal HR, Timpson NJ, et al. International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents. *Circ Cardiovasc Genet*. 2016;9(3):266-278. doi:10.1161/CIRCGENETICS.115.001190.

31. Liu C, Kraja AT, Smith JA, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet*. 2016;48(10):1162-1170. doi:10.1038/ng.3660.

32. Surendran P, Drenos F, Young R, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet*. 2016;48(10):1151-1161. doi:10.1038/ng.3654.

33. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet*. 2017;49(1):54-64. doi:10.1038/ng.3715.

34. Zhu X, Feng T, Tayo BO, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet*. 2015;96(1):21-36. doi:10.1016/j.ajhg.2014.11.011.

35. Fox ER, Young JH, Li Y, et al. Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. *Hum Mol Genet*. 2011;20(11):2273-2284. doi:10.1093/hmg/ddr092.

36. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet*. 2017;49(3):403-415. doi:10.1038/ng.3768.

37. Arora P, Newton-Cheh C. Blood pressure and human genetic variation in the general population. *Curr Opin Cardiol*. 2010;25(3):229-237. doi:10.1097/HCO.0b013e3283383e2c.

38. Nguyen K-DH, Pihur V, Ganesh SK, et al. Effects of rare and common blood pressure gene variants on essential hypertension: results from the Family Blood Pressure Program, CLUE, and Atherosclerosis Risk in Communities studies. *Circ Res*. 2013;112(2):318-326. doi:10.1161/CIRCRESAHA.112.276725.

39. Munroe PB, Barnes MR, Caulfield MJ. Advances in blood pressure genomics. *Circ Res*. 2013;112(10):1365-1379. doi:10.1161/CIRCRESAHA.112.300387.

40. Illumina. datasheet_humanexome_beadchips.pdf. [Internet]. [Accessed: 23 October 2015] Available at: http://products.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_humanexome_beadchips.pdf.).

41. Thorn CF, Ellison DH, Turner ST, Altman RB, Klein TE. PharmGKB summary: Diuretics pathway, pharmacodynamics. *Pharmacogenet Genomics*. 2013;23(8):449-453. doi:10.1097/FPC.0b013e3283636822.

42. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 2008;83(3):311-321. doi:10.1016/j.ajhg.2008.06.024.

43. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011;89(1):82-93. doi:10.1016/j.ajhg.2011.05.029.

44. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892.

45. Ritchie GRS, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods*. 2014;11(3):294-296. doi:10.1038/nmeth.2832.

46. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248.

47. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814.

48. Lee D, Gorkin DU, Baker M, et al. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet*. 2015;47(8):955-961. doi:10.1038/ng.3331.

49. Tomaszewski M, Debiec R, Braund PS, et al. Genetic architecture of ambulatory blood pressure in the general population: insights from cardiovascular gene-centric array. *Hypertens Dallas Tex 1979*. 2010;56(6):1069-1076. doi:10.1161/HYPERTENSIONAHA.110.155721.

50. Exome Chip Design. [Internet]. (2013). [Accessed: 23 October 2015]. Available at: http://genome.sph.umich.edu/wiki/Exome_Chip_Design.

51. Fuchs FD. Why do black Americans have higher prevalence of hypertension?: an enigma still unsolved. *Hypertens Dallas Tex 1979*. 2011;57(3):379-380. doi:10.1161/HYPERTENSIONAHA.110.163196.

52. Zhu X, Young JH, Fox E, et al. Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: contributions from the CARe consortium. *Hum Mol Genet*. 2011;20(11):2285-2295. doi:10.1093/hmg/ddr113.

53. Lifton RP. Molecular genetics of human blood pressure variation. *Science*. 1996;272(5262):676-680.

54. Rotimi CN, Dunston GM, Berg K, et al. In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann Epidemiol*. 2001;11(1):51-58.

55. Sumner AE, Zhou J, Doumatey A, et al. Low HDL-Cholesterol with Normal Triglyceride Levels is the Most Common Lipid Pattern in West Africans and African Americans with Metabolic Syndrome: Implications for Cardiovascular Disease Prevention. *CVD Prev Control*. 2010;5(3):75-80. doi:10.1016/j.cvdpc.2010.07.003.

56. Family Blood Pressure Program. [Internet]. [Accessed: 23 October 2015]. Available at: http://www.biostat.wustl.edu/fbpp/FBPP.shtml.

57. FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertens Dallas Tex 1979*. 2002;39(1):3-9.

58. Howard University – National Human Genome Center. [Internet]. [Accessed: 23 October 2015]. Available at: http://www.genomecenter.howard.edu/units/genetic_epidemiology/default.htm.

59. Adeyemo A, Gerry N, Chen G, et al. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet*. 2009;5(7):e1000564. doi:10.1371/journal.pgen.1000564.

60. Atherosclerosis Risk in Communities. [Internet]. [Accessed: 23 October 2015]. Available at: https://www2.cscc.unc.edu/aric.

61. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702.

62. Cooper R, Rotimi C, Ataman S, et al. The prevalence of hypertension in seven populations of west African origin. *Am J Public Health*. 1997;87(2):160-168.

63. Kang SJ, Chiang CWK, Palmer CD, et al. Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum Mol Genet*. 2010;19(13):2725-2738. doi:10.1093/hmg/ddq154.

64. Ataman SL, Cooper R, Rotimi C, et al. Standardization of blood pressure measurement in an international comparative study. *J Clin Epidemiol*. 1996;49(8):869-877.

65. Cooper R, Puras A, Tracy J, et al. Evaluation of an electronic blood pressure device for epidemiological studies. *Blood Press Monit*. 1997;2(1):35-40.

66. CARDIA Overview. [Internet]. [Accessed: 23 October 2015]. Available at: http://www.cardia.dopm.uab.edu/cardia-overview/overview-more.

67. Friedman GD, Cutter GR, Donahue RP, et al. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol*. 1988;41(11):1105-1116.

68. Grove ML, Yu B, Cochran BJ, et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PloS One*. 2013;8(7):e68095. doi:10.1371/journal.pone.0068095.

69. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795.

70. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

71. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. *Bioinforma Oxf Engl*. 2010;26(22):2867-2873. doi:10.1093/bioinformatics/btq559.

72. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603.

73. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20(1):110-121. doi:10.1101/gr.097857.109.

74. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med*. 2005;24(19):2911-2935. doi:10.1002/sim.2165.

75. Voorman A, Brody J, Chen H, Lumley T. seqMeta: An R package for meta-analyzing region-based tests of rare DNA variants. 2013: http://cran.r-project.org/web/packages/seqMeta/.

76. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91(2):224-237. doi:10.1016/j.ajhg.2012.06.007.

77. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*. 2010;11(7):459-463. doi:10.1038/nrg2813.

78. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661-678. doi:10.1038/nature05911.

79. Pirie A, Wood A, Lush M, Tyrer J, Pharoah PDP. The effect of rare variants on inflation of the test statistics in case-control analyses. *BMC Bioinformatics*. 2015;16:53. doi:10.1186/s12859-015-0496-1.

80. Ackerman C, Locke AE, Feingold E, et al. An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am J Hum Genet*. 2012;91(4):646-659. doi:10.1016/j.ajhg.2012.08.017.

81. Löffler M, Morote-Garcia JC, Eltzschig SA, Coe IR, Eltzschig HK. Physiological roles of vascular nucleoside transporters. *Arterioscler Thromb Vasc Biol*. 2007;27(5):1004-1013. doi:10.1161/ATVBAHA.106.126714.

82. Ritzel MW, Ng AM, Yao SY, et al. Molecular identification and characterization of novel human and mouse concentrative Na+-nucleoside cotransporter proteins (hCNT3 and mCNT3) broadly selective for purine and pyrimidine nucleosides (system cib). *J Biol Chem*. 2001;276(4):2914-2927. doi:10.1074/jbc.M007746200.

83. Horswell SD, Fryer LGD, Hutchison CE, et al. CDKN2B expression in adipose tissue of familial combined hyperlipidemia patients. *J Lipid Res*. 2013;54(12):3491-3505. doi:10.1194/jlr.M041814.

84. LaDisa JF, Bozdag S, Olson J, Ramchandran R, Kersten JR, Eddinger TJ. Gene Expression in Experimental Aortic Coarctation and Repair: Candidate Genes for Therapeutic Intervention? *PloS One*. 2015;10(7):e0133356. doi:10.1371/journal.pone.0133356.

85. Ye X, Zhang HM, Qiu Y, et al. Coxsackievirus-induced miR-21 disrupts cardiomyocyte interactions via the downregulation of intercalated disk components. *PLoS Pathog*. 2014;10(4):e1004070. doi:10.1371/journal.ppat.1004070.

86. Staples CJ, Myers KN, Beveridge RDD, et al. Ccdc13 is a novel human centriolar satellite protein required for ciliogenesis and genome stability. *J Cell Sci*. 2014;127(Pt 13):2910-2919. doi:10.1242/jcs.147785.

87. Nauli SM, Jin X, Hierck BP. The mechanosensory role of primary cilia in vascular hypertension. *Int J Vasc Med*. 2011;2011:376281. doi:10.1155/2011/376281.

88. Rahbari-Oskoui F, Williams O, Chapman A. Mechanisms and management of hypertension in autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc*. 2014;29(12):2194-2201. doi:10.1093/ndt/gft513.

89. Borges BE, Appel MH, Cofré AR, et al. The flavo-oxidase QSOX1 supports vascular smooth muscle cell migration and proliferation: Evidence for a role in neointima growth. *Biochim Biophys Acta*. 2015;1852(7):1334-1346. doi:10.1016/j.bbadis.2015.03.002.

90. de Andrade CR, Stolf BS, Debbas V, et al. Quiescin sulfhydryl oxidase (QSOX) is expressed in the human atheroma core: possible role in apoptosis. *In Vitro Cell Dev Biol Anim*. 2011;47(10):716-727. doi:10.1007/s11626-011-9461-0.

91. Mebazaa A, Vanpoucke G, Thomas G, et al. Unbiased plasma proteomics for novel diagnostic biomarkers in cardiovascular disease: identification of quiescin Q6 as a candidate biomarker of acutely decompensated heart failure. *Eur Heart J*. 2012;33(18):2317-2324. doi:10.1093/eurheartj/ehs162.

92. Doehner W. Diagnostic biomarkers in cardiovascular disease: the proteomics approach. *Eur Heart J*. 2012;33(18):2249-2251. doi:10.1093/eurheartj/ehs187.

93. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014;506(7487):185-190. doi:10.1038/nature12975.

94. Richards AL, Leonenko G, Walters JT, et al. Exome arrays capture polygenic rare variant contributions to schizophrenia. *Hum Mol Genet*. 2016;25(5):1001-1007. doi:10.1093/hmg/ddv620.

95. Turner, SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv. 2014. DOI: 10.1101/005165.

96. CDC. National Center for Health Statistics – Hypertension. http://www.cdc.gov/nchs/fastats/hypertension.htm. Updated April 27, 2016.

97. Rudin A. Bartter's syndrome. A review of 28 patients followed for 10 years. *Acta Med Scand*. 1988;224(2):165-171.

98. Simon DB, Nelson-Williams C, Bia MJ, et al. Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na-Cl cotransporter. *Nat Genet*. 1996;12(1):24-30. doi:10.1038/ng0196-24.

99. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73-80. doi:10.1161/CIRCGENETICS.108.829747.

100. Yu B, Pulit SL, Hwang S-J, et al. Rare Exome Sequence Variants in CLCN6 Reduce Blood Pressure Levels and Hypertension Risk. *Circ Cardiovasc Genet*. 2016;9(1):64-70. doi:10.1161/CIRCGENETICS.115.001215.

101. Morrison AC, Voorman A, Johnson AD, et al. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet*. 2013;45(8):899-901. doi:10.1038/ng.2671.

102. Morrison AC, Huang Z, Yu B, et al. Practical Approaches for Whole-Genome Sequence Analysis of Heart- and Blood-Related Traits. *Am J Hum Genet*. 2017;100(2):205-215. doi:10.1016/j.ajhg.2016.12.009.

103. Levy D, DeStefano AL, Larson MG, et al. Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertens Dallas Tex 1979*. 2000;36(4):477-483.

104. Franklin SS, Gustin W, Wong ND, et al. Hemodynamic patterns of age-related changes in blood pressure. The Framingham Heart Study. *Circulation*. 1997;96(1):308-315.

105. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinforma Oxf Engl*. 2007;23(21):2947-2948. doi:10.1093/bioinformatics/btm404.

106. Li W, Cowley A, Uludag M, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*. 2015;43(W1):W580-W584. doi:10.1093/nar/gkv279.

107. Thomas PD, Kejariwal A, Guo N, et al. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res*. 2006;34(Web Server issue):W645-W650.

108. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20. doi:10.1002/0471142905.hg0720s76.

109. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinforma Oxf Engl*. 2014;30(6):884-886. doi:10.1093/bioinformatics/btt607.

110. Fang L, Li D, Welling PA. Hypertension resistance polymorphisms in ROMK (Kir1.1) alter channel function by different mechanisms. *Am J Physiol Renal Physiol*. 2010;299(6):F1359-F1364. doi:10.1152/ajprenal.00257.2010.

111. Monette MY, Rinehart J, Lifton RP, Forbush B. Rare mutations in the human Na-K-Cl cotransporter (NKCC2) associated with lower blood pressure exhibit impaired processing and transport function. *Am J Physiol Renal Physiol*. 2011;300(4):F840-F847. doi:10.1152/ajprenal.00552.2010.

112. Acuña R, Martínez-de-la-Maza L, Ponce-Coria J, et al. Rare mutations in SLC12A1 and SLC12A3 protect against hypertension by reducing the activity of renal salt cotransporters. *J Hypertens*. 2011;29(3):475-483. doi:10.1097/HJH.0b013e328341d0fd.

113. Ajay SS, Parker SCJ, Abaan HO, Fajardo KVF, Margulies EH. Accurate and comprehensive sequencing of personal genomes. *Genome Res*. 2011;21(9):1498-1505. doi:10.1101/gr.123638.111.

114. Clark MJ, Chen R, Lam HYK, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol*. 2011;29(10):908-914. doi:10.1038/nbt.1975.

115. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*. 2014;15:247. doi:10.1186/1471-2105-15-247.

116. Fuentes Fajardo KV, Adams D, NISC Comparative Sequencing Program, et al. Detecting false-positive signals in exome sequencing. *Hum Mutat*. 2012;33(4):609-613. doi:10.1002/humu.22033.

117. Huang X-F, Wu J, Lv J-N, Zhang X, Jin Z-B. Identification of false-negative mutations missed by next-generation sequencing in retinitis pigmentosa patients: a

complementary approach to clinical genetic diagnostic testing. *Genet Med Off J Am Coll Med Genet*. 2015;17(4):307-311. doi:10.1038/gim.2014.193.

118. Guyton AC, Coleman TG, Cowley AV, Scheel KW, Manning RD, Norman RA. Arterial pressure regulation. Overriding dominance of the kidneys in long-term regulation and in hypertension. *Am J Med*. 1972;52(5):584-594.

119. Hall JE. Control of blood pressure by the renin-angiotensin-aldosterone system. *Clin Cardiol*. 1991;14(8 Suppl 4):IV6-IV21; discussion IV51-IV55.

120. Carey RM. The intrarenal renin-angiotensin system in hypertension. *Adv Chronic Kidney Dis*. 2015;22(3):204-210. doi:10.1053/j.ackd.2014.11.004.

121. Paul M, Poyan Mehr A, Kreutz R. Physiology of local renin-angiotensin systems. *Physiol Rev*. 2006;86(3):747-803. doi:10.1152/physrev.00036.2005.

122. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888.

123. Hulur I, Gamazon ER, Skol AD, et al. Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics*. 2015;16:138. doi:10.1186/s12864-015-1292-z.

124. Walsh AM, Whitaker JW, Huang CC, et al. Integrative genomic deconvolution of rheumatoid arthritis GWAS loci into gene and cell type associations. *Genome Biol*. 2016;17:79. doi:10.1186/s13059-016-0948-6.

125. Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*. 2015;200(4):1285-1295. doi:10.1534/genetics.115.178616.

126. Kvale MN, Hesselson S, Hoffmann TJ, et al. Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*. 2015;200(4):1051-1060. doi:10.1534/genetics.115.178905.

127. Barbeira A, Shah KP, Torres JM, Wheeler HE, Torstenson ES, Edwards T, et al. MetaXcan: Sum- 537 mary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results. bioRxiv. 538 2016 mar;Available from: http://biorxiv.org/content/early/2016/03/23/045260.abstract.

128. Maruthur NM, Li M, Halushka MK, et al. Genetics of Plasma Soluble Receptor for Advanced Glycation End-Products and Cardiovascular Outcomes in a Community-based Population: Results from the Atherosclerosis Risk in Communities Study. *PloS One*. 2015;10(6):e0128452. doi:10.1371/journal.pone.0128452.

129. Li M, Maruthur NM, Loomis SJ, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. *Sci Rep*. 2017;7(1):2812. doi:10.1038/s41598-017-02287-x.

130. Kapoor A, Bakshy K, Xu L, et al. Rare coding TTN variants are associated with electrocardiographic QT interval in the general population. *Sci Rep*. 2016;6:28356. doi:10.1038/srep28356.

131. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-660. doi:10.1126/science.1262110.

132. Zhang X, Gierman HJ, Levy D, et al. Synthesis of 53 tissue and cell line expression QTL datasets reveals master eQTLs. *BMC Genomics*. 2014;15:532. doi:10.1186/1471-2164-15-532.

133. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. doi:10.1038/nprot.2008.211.

134. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13. doi:10.1093/nar/gkn923.

135. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9):1760-1774. doi:10.1101/gr.135350.111.

136. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404.

137. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352.

138. Picard Tools webpage. [http://picard.sourceforge.net].

139. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma Oxf Engl*. 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033.

140. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012;22(9):1813-1831. doi:10.1101/gr.136184.111.

141. Strino F, Lappe M. Identifying peaks in *-seq data using shape information. *BMC Bioinformatics*. 2016;17 Suppl 5:206. doi:10.1186/s12859-016-1042-5.

142. Ghandi M, Lee D, Mohammad-Noori M, Beer MA. Enhanced regulatory sequence prediction using gapped k-mer features. *PLoS Comput Biol*. 2014;10(7):e1003711. doi:10.1371/journal.pcbi.1003711.

143. Lee D, Karchin R, Beer MA. Discriminative prediction of mammalian enhancers from DNA sequence. *Genome Res*. 2011;21(12):2167-2180. doi:10.1101/gr.121905.111.

144. Gorkin DU, Lee D, Reed X, et al. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Res*. 2012;22(11):2290-2301. doi:10.1101/gr.139360.112.

145. Lee D. LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinforma Oxf Engl*. 2016;32(14):2196-2198. doi:10.1093/bioinformatics/btw142.

146. Seunggeun Lee, with contributions from Larisa Miropolsky and Michael Wu (2015). SKAT: SNP-Set (Sequence) Kernel Association Test. R package version 1.1.2. https://CRAN.R-project.org/package=SKAT.

147. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47(9):1091-1098. doi:10.1038/ng.3367.

148. Postema PG, Wilde AAM. The measurement of the QT interval. *Curr Cardiol Rev*. 2014;10(3):287-294.

149. Dekker JM, Crow RS, Hannan PJ, Schouten EG, Folsom AR, ARIC Study. Heart rate-corrected QT interval prolongation predicts risk of coronary heart disease in black and white middle-aged men and women: the ARIC study. *J Am Coll Cardiol*. 2004;43(4):565-571. doi:10.1016/j.jacc.2003.09.040.

150. Newton-Cheh C, Larson MG, Corey DC, et al. QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: The Framingham Heart Study. *Heart Rhythm*. 2005;2(3):277-284. doi:10.1016/j.hrthm.2004.11.009.

151. Busjahn A, Knoblauch H, Faulhaber HD, et al. QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation*. 1999;99(24):3161-3164.

152. Arking DE, Pfeufer A, Post W, et al. A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet*. 2006;38(6):644-651. doi:10.1038/ng1790.

153. Tomás M, Napolitano C, De Giuli L, et al. Polymorphisms in the NOS1AP gene modulate QT interval duration and risk of arrhythmias in the long QT syndrome. *J Am Coll Cardiol*. 2010;55(24):2745-2752. doi:10.1016/j.jacc.2009.12.065.

154. Avery CL, Wassel CL, Richard MA, et al. Fine mapping of QT interval regions in global populations refines previously identified QT interval loci and identifies signals unique to African and Hispanic descent populations. *Heart Rhythm*. 2017;14(4):572-580. doi:10.1016/j.hrthm.2016.12.021.

155. Arking DE, Pulit SL, Crotti L, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet*. 2014;46(8):826-836. doi:10.1038/ng.3014.

156. Li Q, Gao T, Yuan Y, et al. Association of CYP17A1 Genetic Polymorphisms and Susceptibility to Essential Hypertension in the Southwest Han Chinese Population. *Med Sci Monit Int Med J Exp Clin Res*. 2017;23:2488-2499.

157. Li C, Kim YK, Dorajoo R, et al. Genome-Wide Association Study Meta-Analysis of Long-Term Average Blood Pressure in East Asians. *Circ Cardiovasc Genet*. 2017;10(2):e001527. doi:10.1161/CIRCGENETICS.116.001527.

158. Lin CG, Chen C-C, Leu S-J, Grzeszkiewicz TM, Lau LF. Integrin-dependent functions of the angiogenic inducer NOV (CCN3): implication in wound healing. *J Biol Chem*. 2005;280(9):8229-8237. doi:10.1074/jbc.M404903200.

159. Lin CG, Leu S-J, Chen N, et al. CCN3 (NOV) is a novel angiogenic regulator of the CCN protein family. *J Biol Chem*. 2003;278(26):24200-24208. doi:10.1074/jbc.M302028200.

160. Shimoyama T, Hiraoka S, Takemoto M, et al. CCN3 inhibits neointimal hyperplasia through modulation of smooth muscle cell growth and migration. *Arterioscler Thromb Vasc Biol*. 2010;30(4):675-682. doi:10.1161/ATVBAHA.110.203356.

161. Guo D-C, Grove ML, Prakash SK, et al. Genetic Variants in LRP1 and ULK4 Are Associated with Acute Aortic Dissections. *Am J Hum Genet*. 2016;99(3):762-769. doi:10.1016/j.ajhg.2016.06.034.

162. Niu W-Q, You Y-G, Qi Y. Strong association of methylenetetrahydrofolate reductase gene C677T polymorphism with hypertension and hypertension-in-pregnancy in Chinese: a meta-analysis. *J Hum Hypertens*. 2012;26(4):259-267. doi:10.1038/jhh.2011.11.

163. Frosst P, Blom HJ, Milos R, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet*. 1995;10(1):111-113. doi:10.1038/ng0595-111.

164. Nakata Y, Katsuya T, Takami S, et al. Methylenetetrahydrofolate reductase gene polymorphism: relation to blood pressure and cerebrovascular disease. *Am J Hypertens*. 1998;11(8 Pt 1):1019-1023.

165. Kosmas IP, Tatsioni A, Ioannidis JPA. Association of C677T polymorphism in the methylenetetrahydrofolate reductase gene with hypertension in pregnancy and pre-eclampsia: a meta-analysis. *J Hypertens*. 2004;22(9):1655-1662.

166. Van Woudenberg M, Shin J, Bernard M, et al. CYP17A1 and Blood Pressure Reactivity to Stress in Adolescence. *Int J Hypertens*. 2015;2015:734586. doi:10.1155/2015/734586.

167. Airik R, Slaats GG, Guo Z, et al. Renal-retinal ciliopathy gene Sdccag8 regulates DNA damage response signaling. *J Am Soc Nephrol JASN*. 2014;25(11):2573-2583. doi:10.1681/ASN.2013050565.

168. Laflamme N, Leblanc JF, Mailloux J, Faure N, Labrie F, Simard J. Mutation R96W in cytochrome P450c17 gene causes combined 17 alpha-hydroxylase/17-20-lyase deficiency in two French Canadian patients. *J Clin Endocrinol Metab*. 1996;81(1):264-268. doi:10.1210/jcem.81.1.8550762.

169. Goldsmith O, Solomon DH, Horton R. Hypogonadism and mineralocorticoid excess. The 17-hydroxylase deficiency syndrome. *N Engl J Med*. 1967;277(13):673-677. doi:10.1056/NEJM196709282771302.

170. Imai T, Yanase T, Waterman MR, Simpson ER, Pratt JJ. Canadian Mennonites and individuals residing in the Friesland region of The Netherlands share the same molecular basis of 17 alpha-hydroxylase deficiency. *Hum Genet*. 1992;89(1):95-96.

171. Camats N, Üstyol A, Atabek ME, Dick B, Flück CE. A novel CYP17A1 deletion causes a functional knockout of the steroid enzyme 17-hydroxylase and 17,20-lyase in a Turkish family and illustrates the precise role of the CYP17A1 gene. *Clin Case Rep*. 2015;3(10):793-797. doi:10.1002/ccr3.343.

172. Grimaldi F. [Adrenal tumors as a cause of hypertension]. *G Ital Nefrol Organo Uff Della Soc Ital Nefrol*. 2010;27 Suppl 50:S86-S90.

173. Chappell MC. Nonclassical renin-angiotensin system and renal function. *Compr Physiol*. 2012;2(4):2733-2752. doi:10.1002/cphy.c120002.

174. Freedman BI, Cohen AH. Hypertension-attributed nephropathy: what's in a name? *Nat Rev Nephrol*. 2016;12(1):27-36. doi:10.1038/nrneph.2015.172.

175. Kopp JB. Rethinking hypertensive kidney disease: arterionephrosclerosis as a genetic, metabolic, and inflammatory disorder. *Curr Opin Nephrol Hypertens*. 2013;22(3):266-272. doi:10.1097/MNH.0b013e3283600f8c.

176. Ehret G. Next Steps for Gene Identification in Primary Hypertension Genomics. *Hypertens Dallas Tex 1979*. 2017;70(4):695-697. doi:10.1161/HYPERTENSIONAHA.117.09719.

Permissions

Chapter 2 is reprinted from the final accepted preprint of the journal article: Nandakumar P, Lee D, Richard MA, Tekola-Ayele F, Tayo BO, Ware E, Sung YJ, Salako B, Ogunniyi A, Gu CC, Grove ML, Fornage M, Kardia S, Rotimi C, Cooper RS, Morrison AC, Ehret GB, Chakravarti A. Rare coding variants associated with blood pressure variation in 15,914 individuals of African ancestry. *J Hypertens*. 2017;Vol 35; Issue 7; 1381–1389. doi: 10.1097/HJH.0000000000001319. http://journals.lww.com/jhypertension/Abstract/2017/07000/Rare_coding_variants_associated_with_blood.10.aspx

Copyright Clearance Center RightsLink®

Home | Create Account | Help

Wolters Kluwer

Title: Rare coding variants associated with blood pressure variation in 15914 individuals of African ancestry

Author: Priyanka Nandakumar, Dongwon Lee, Melissa Richard, et al

Publication: Journal of Hypertension

Publisher: Wolters Kluwer Health, Inc.

Date: Aug 17, 0702

Copyright © 2017, Copyright (C) 2017 Wolters Kluwer Health, Inc. All rights reserved.

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

Quick Price Estimate

Permisssion to use a full article in a dissertation/thesis can only be granted to the author of the original Wolters Kluwer content. You are permitted to reuse a specific portion of the article such as a figure, abstract, or excerpt.

I would like to... reuse in a dissertation/thesis

This service provides permission for reuse only.

Select your currency USD - $

Requestor Type Individual

Portion Full article

Number of pages requested 9

Are you the author of this Wolters Kluwer article? Yes

If you do not already have a copy of the content you wish to reuse, you can purchase the article through journals.lww.com and extract the images.

Author of this article? Select yes only if you are also re-using the material into your own publication.

Quick Price Click Quick Price

QUICK PRICE | CONTINUE

To request permission for a type of use not listed, please email healthpermissions@wolterskluwer.com.

Copyright © 2017 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Certain sections (designated text, Table 1, Figure 1, and GERA, ICBP, and UBK acknowledgements) in Chapter 4 are reprinted from the final accepted preprint of the final journal article:
Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, Iribarren C, Chakravarti A, Risch N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature genetics*. 2017;49(1):54-64. doi:10.1038/ng.3715. (permissions on next page).

The legend for Figure 1 in Chapter 4 was printed in the original article text with an error (stating 51 tissues instead of 45) and is the process of correction; this reprint includes the corrected value.

Curriculum Vitae

Priyanka Nandakumar
Email: pnandak1@jhmi.edu
Baltimore, MD

EDUCATION

- **PhD Candidate at Johns Hopkins University, School of Medicine** 2017
  - Human Genetics and Molecular Biology
  - Dissertation title: *Identifying the genes and tissues regulating blood pressure and its variation*

- **Carnegie Mellon University, Mellon College of Science** 2012
  - B.S. Biological Sciences, minor in chemistry (Honors)
  - Cumulative GPA: 4.00/4.00

TECHNICAL SKILLS AND PROFESSIONAL DEVELOPMENT

- **Programming**: Experienced with R, Linux/shell scripting, and high performance computing; previous use of python, Perl, Java, and C++ from graduate and undergraduate research and coursework.
- **Genomics Tools and Skills**: Genetic association/NGS analysis including DNA and RNA sequencing pipeline tools and file formats, R/Bioconductor packages, meta-analysis, genome browsers, basic statistics.
- **Professional Development**: Johns Hopkins Entrepreneurial Bootcamp - learned about commercializing biomedical products and strategies created in the laboratory; worked collaboratively on a team, which was selected as one of four teams for a "Shark Tank" oral presentation out of 20+ teams at the end of the course
-

DISSERTATION RESEARCH EXPERIENCE

- **Graduate student (Aravinda Chakravarti lab)** 05/2013 – Present
  Institute of Genetic Medicine, Johns Hopkins School of Medicine Baltimore, MD
  - Previously demonstrated an enrichment of blood pressure loci in artery eQTLs; current work follows up with BP-tissue-specific gene identification
  - Analysis of sequence and array data in blood pressure and with seqMeta, PLINK, R and other genomics tools and software
  - Case-control analysis of GWAS data in Hirschsprung disease
  - RNA-seq analysis (mRNA & miRNA) from raw data through differential expression analysis
  - Experienced with intra- and inter-laboratory and institutional collaborations

TEACHING EXPERIENCE

Graduate Teaching Assistant 11/2014 – 12/2014

Course: Fundamentals of Genetics Baltimore, MD
Johns Hopkins School of Medicine
- Human Genetics section; office hours, designing and grading of problem sets and exam questions

· **Undergraduate Teaching Assistant** 08/2010 – 12/2010
Course: Physics I for Science Students Pittsburgh, PA
Department of Physics, Carnegie Mellon University
- Assist students with exercises and homework

WORK AND OUTREACH EXPERIENCE

· **Science Outreach Program Coordinator and Member** 05/2013 – 12/2014
Johns Hopkins School of Medicine Baltimore, MD
- Collaboration with a nonprofit organization (Village Learning Place) in running a six-week summer science enrichment program for children in grades 3-6
- Worked on a team to manage and participate in the program by preparing curricula, teaching, and interfacing with the organization

Outreach Mentor 09/2009 – 03/2012
· Departments of Biological Sciences & Physics, Pittsburgh, PA
Carnegie Mellon University
- In charge of two AP Biology students for one day outreach events in which the students perform experiments
- Designed & assisted 8th grade student with experiment for Pennsylvania Junior Academy of Science Fair

AWARDS, HONORS AND AFFILIATIONS

Graduate:
- 2014 Winner, Johns Hopkins MD-GEM Genetic Epidemiology internal mini-grant (Wolfe St. Competition); a grant submitted as a partnership between a pre-doctoral student and a postdoctoral candidate (Adrienne Tin), titled "Differential Transcriptome Profiling of African Americans with and without both Hypertension and Chronic Kidney Disease"
- Affiliations: American Society of Human Genetics (2015-2016), American Heart Association (2016)
- Reviewed a *Scientific Reports* article (2016) on a genetic study of Hirschsprung disease

· Undergraduate:
- Honor societies: Phi Beta Kappa, Phi Kappa Phi
- University Honors & College Research Honors

PUBLICATIONS

Published (* designates equal contributions)
Nandakumar P, Lee D, Richard MA, Tekola-Ayele F, Tayo BO, Ware E, Sung YJ, Salako B, Ogunniyi A, Gu CC, Grove ML, Fornage M, Kardia S, Rotimi C,

Cooper RS, Morrison AC, Ehret GB, Chakravarti A. Rare coding variants associated with blood pressure variation in 15,914 individuals of African ancestry. *J Hypertens*. 2017 Feb 22. doi: 10.1097/HJH.0000000000001319.

Nandakumar P*, Tin A*, Grove ML, Ma J, Boerwinkle E, Coresh J, Chakravarti A. MicroRNAs in the miR-17 and miR-15 families are downregulated in chronic kidney disease with hypertension. Endlich N, ed. PLoS ONE. 2017;12(8):e0176734. doi:10.1371/journal.pone.0176734.

Wain LV, Vaez A, Jansen R, Joehanes R, van der Most PJ, Erzurumluoglu AM et al. Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. Hypertension. 2017 Jul 24. pii: HYPERTENSIONAHA.117.09438. doi: 0.1161/HYPERTENSIONAHA.117.09438.

He KY, Wang H, Cade BE, **Nandakumar P**, Giri A, Ware EB, Haessler J, Liang J, Smith JA, Franceschini N9, Le TH, Kooperberg C, Edwards TL, Kardia SL, Lin X, Chakravarti A, Redline S, Zhu X. Rare variants in fox-1 homolog A (RBFOX1) are associated with lower blood pressure. *PLoS Genet*. 2017 Mar 27;13(3):e1006678.

Hoffmann TJ, Ehret GB, **Nandakumar P**, Ranatunga D, Schaefer C, Kwok PY, Iribarren C, Chakravarti A, Risch N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet*. 2017 Jan;49(1):54-64. doi: 10.1038/ng.3715.

Tang CS, Gui H, Kapoor A, Kim JH, Luzón-Toro B, Pelet A, Burzynski G, Lantieri F, So MT, Berrios C, Shin HD, Fernández RM, Le TL, Verheij JB, Matera I, Cherny SS, **Nandakumar P**, Cheong HS, Antiñolo G, Amiel J, Seo JM, Kim DY, Oh JT, Lyonnet S, Borrego S, Ceccherini I, Hofstra RM, Chakravarti A, Kim HY, Sham PC, Tam PK, Garcia-Barceló MM. Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease. *Hum Mol Genet*. 2016 Oct 3. pii: ddw333. doi: 10.1093/hmg/ddw333. [Epub ahead of print]

Kapoor A, Bakshy K, Xu L, **Nandakumar P**, Lee D, Boerwinkle E, Grove ML, Arking DE, Chakravarti A. Rare coding TTN variants are associated with electrocardiographic QT interval in the general population. *Sci Rep*. 2016 Jun 20;6:28356. doi: 10.1038/srep28356.

Sung YJ, Basson J, Cheng N, Nguyen KDH, **Nandakumar P**, Hunt SC, Arnett DK, Dávila-Román V, G, Rao DC, Chakravarti A, The Role of Rare Variants in Systolic Blood Pressure: Analysis of ExomeChip Data in HyperGEN African Americans. *Hum Hered* 2015;79:20-27.

Grahnen, J. A., **Nandakumar, P.**, Kubelka, J., & Liberles, D. A. (2011). Biophysical and structural considerations for protein sequence evolution. *BMC Evolutionary Biology*, 11, 361.

Under review, submitted or in prep (* designates equal contributions)
Nandakumar P, Morrison AC, Grove ML, Boerwinkle E, Chakravarti A. Contributions of Rare Coding Variants in Hypotension Syndrome Genes to Population Blood Pressure Variation. (in prep)

Chatterjee S*, **Nandakumar P***, Auer D, Gabriel S, Chakravarti A. The gastrointestinal 'parts list': characterizing embryonic gut development by temporal transcriptional profiling of wildtype and Ret-deficient mice. (in prep)

Wang H, **Nandakumar P**, Tekola-Ayele F, Tayo BO, Ware E, Gu CC, Fornage M, Kardia S, Rotimi C, Cooper RS, Rao DC, Morrison AC, Ehret GB, Chakravarti A, Zhu X. Rare variants within multiple genes at a chromosome 1q locus contributes to blood pressure variation: the combined use of linkage-association tests. (in prep)

Fadista J, Lund M, Geller F, **Nandakumar P**, Chatterjee S, Matsson H, Granstrom AL, Wester T, Salo P, Virtanen V, Carstensen L, Bybjerg-Grauholm J, Hougaard DM, Pakarinen M, Perola M, Nordenskjöld A, Chakravarti A, Melbye M, Feenstra B. GWAS of Danish Hirschsprung disease detects a novel low frequency associated variant and makes case/control prediction with high accuracy. (under review)

PRESENTATIONS (* designates equal contributions)

Nandakumar P, Lee D, Hoffmann TJ, Ehret GB, Ranatunga D, Schaefer C, Kwok PY, Iribarren C, Risch N, Chakravarti A. Analysis of cis-regulatory elements in tissues with potential relevance to blood pressure regulation. Poster presentation at Gordon Research Conference Human Genetics & Genomics. 2017.

Nandakumar P, Lee D, Richard MA, Tekola-Ayele F, Tayo BO, Ware E, Sung YJ, Salako B, Ogunniyi A, Gu CC, Grove ML, Fornage M, Kardia S, Rotimi C, Cooper RS, Morrison AC, Ehret GB, Chakravarti A. Rare coding variants associated with blood pressure in 15,914 individuals of African ancestry. Poster presentation at American Society of Human Genetics, October 21, 2016.

Nandakumar P, Lee D, Richard, Melissa A., Tekola-Ayele, F, Tayo, B, Ware, E, Gu,CC, Fornage M, Kardia S, Rotimi C, Cooper R, Rao DC, Morrison A, Ehret G, Chakravarti A. Rare coding variants associated with blood pressure in 14,620 individuals of African ancestry. Poster presentation at American Society of Human Genetics, October 7, 2015.

Nandakumar, P* and Tin, A*. Differential Expression Analysis of Hypertension and Chronic Kidney Disease (CKD) in African Americans. Oral Presentation for the Johns Hopkins MD-GEM seminar series. May 3, 2016.

Nandakumar, P*, Tin, A* (presenter), Boerwinkle E, Grove ML, Coresh J, Chakravarti A. Differential Transcriptome Profiling of African Americans with Uncontrolled Hypertension and Chronic Kidney Disease (CKD) versus Controlled Hypertension and without CKD: Study Design. Poster Presentation at Genetics Research Day 2016 Johns Hopkins MD-GEM. (February 2016)

Nandakumar, P*, Tin, A*, Boerwinkle E, Grove ML, Coresh J, Chakravarti A. Differential Transcriptome Profiling of African Americans with Uncontrolled Hypertension and Chronic Kidney Disease (CKD) versus Controlled Hypertension and without CKD: Study Design. Poster Presentation at Genetics Research Day 2015 Johns Hopkins MD-GEM. (February 2015)